



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 119997

TO: Michael Borin
Location: rem/2a55/2c70
Art Unit: 1631
Friday, April 23, 2004

Case Serial Number: 09/528682

From: Noble Jarrell
Location: Biotech-Chem Library
Rem 1B71
Phone: 272-2556

Noble.jarrell@uspto.gov

Search Notes

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:48:31 ; Search time 11.6667 Seconds
(without alignments)
131.920 Million cell updates/sec

Title: US-09-528-682-1_COPY_64_79
Perfect score: 16
Sequence: 1 LSLRSAHLAQSSILSG 16

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 78.*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	100.0	258	1 QLECA	heat-labile entero
2	7	43.8	258	1 XVVCA	cholera enterotoxi
3	7	43.8	1333	2 S63403	probable membrane
4	6	37.5	86	2 T03209	hypothetical prote
5	6	37.5	120	2 B83415	hypothetical prote
6	6	37.5	123	2 S44905	ribosomal protein
7	6	37.5	128	2 D72750	hypothetical prote
8	6	37.5	140	2 S72252	hemoglobin, extrac
9	6	37.5	207	1 Q08CU2	hypothetical 21.9K
10	6	37.5	207	2 AG2274	methyltransferase
11	6	37.5	232	2 F82729	conserved hypothet
12	6	37.5	241	2 AF3327	transcription regu
13	6	37.5	245	2 S12089	hypothetical prote
14	6	37.5	270	2 C75257	hypothetical prote
15	6	37.5	294	2 D90578	elongation factor
16	6	37.5	344	2 T26323	hypothetical prote
17	6	37.5	345	2 T36989	probable transcrip
18	6	37.5	377	2 AC1723	hypothetical prote
19	6	37.5	377	2 AH1352	hypothetical prote
20	6	37.5	395	2 AG3233	hypothetical prote
21	6	37.5	398	2 T27254	hypothetical prote
22	6	37.5	462	2 H75438	3-phosphoshikimate
23	6	37.5	469	2 E75423	two-component sens
24	6	37.5	486	2 AE2072	hypothetical prote
25	6	37.5	512	2 H90010	PTS system, nannit
26	6	37.5	518	1 S22385	phosphotransferase
27	6	37.5	537	2 AF2785	lipoprotein (impor
28	6	37.5	550	2 E90723	probable fumarate
29	6	37.5	550	2 E85574	probable fumarate

43K antigen (AF157
hypothetical prote
hypothetical prote
probable oxidoredu
hypothetical prote
protein P11N15.9 [1
MNE1 protein - yea
hypothetical prote
SITS-binding prote
beta-adrenergic-re
probable membrane
ATP-dependent heli
Mg2+-transporting
ionotropic glutama
probable SNF2 subf
SNQ2 protein - Yea

30 6 37.5 562 2 H97564
31 6 37.5 565 2 T02899
32 6 37.5 576 1 S62534
33 6 37.5 588 2 B95406
34 6 37.5 590 2 T20153
35 6 37.5 655 2 G96524
36 6 37.5 663 2 S67259
37 6 37.5 696 2 H83024
38 6 37.5 697 1 S04987
39 6 37.5 700 1 A41615
40 6 37.5 737 2 S55117
41 6 37.5 822 2 E75523
42 6 37.5 908 2 B39083
43 6 37.5 950 2 T51134
44 6 37.5 1339 2 A84683
45 6 37.5 1501 2 S50992

ALIGNMENTS

RESULT 1
QLECA
heat-labile enterotoxin A precursor - Escherichia coli
C:Species: Escherichia coli
C:Date: 30-Apr-1981 #sequence revision 17-Oct-1997 #text_change 18-Jun-1999
C:Accession: I55231; A01817; A26946
R:Yamamoto, T.; Tamura, T.; Yokota, T.
J. Biol. Chem. 259, 5037-5044, 1984
A:Title: Primary structure of heat-labile enterotoxin produced by Escherichia coli pathc
A:Reference number: I55231; MUID:84185610; PMID:6325417
A:Accession: I55231
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-258 <RES>
A:Cross-references: GB:X01995; NID:9148027; PIDN:AAA24685.1; PID:G148028
R:Spicer, E.K.; Noble, J.A.
J. Biol. Chem. 257, 5716-5721, 1982
A:Title: Escherichia coli heat-labile enterotoxin. Nucleotide sequence of the A subunit
A:Reference number: A01817; MUID:82167425; PMID:6279611
A:Accession: A01817
A:Molecule type: DNA
A:Residues: 1-21, 'R', 23-36, 'FRS', 40-44, 46-92, 'Y', 94-99, 'LTYI', 105-107, 111-118, 'IS', 121-
A:Cross-references: EMBL:V00275; NID:941339; PIDN:CAA23532.1; PID:G41340
A:Note: the authors translated the codon TAT for residue 93 as Ser
R:Yamamoto, T.; Gojohori, T.; Yokota, T.
J. Bacteriol. 169, 1352-1357, 1987
A:Title: Evolutionary origin of pathogenic determinants in enterotoxigenic Escherichia c
A:Reference number: A26946; MUID:87137303; PMID:3546273
A:Accession: A26946
A:Molecule type: DNA
A:Residues: 1-21, 'R', 23-206, 'N', 208-230, 'E', 232-255, 'D', 257-258 <YAM>
A:Cross-references: EMBL:M15363
C:Comment: The heat-labile enterotoxin molecule contains one A chain and five or six B c
clase.
C:Genetics:
A:Gene: eltA
C:Superfamily: heat-labile enterotoxin chain A
C:Keywords: enterotoxin
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-258/Product: heat-labile enterotoxin chain A #status predicted <MAT>

Query Match 100.0%; Score 16; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 8.3e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSAHLAQSSILSG 16
|||
Db 82 LSLRSAHLAQSSILSG 97

RESULT 2
XVVCA

R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dracoli, I.; Sellers,
 I. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A;Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
 A;Reference number: A82035; MUID:20406833; PMID:10952301
 A;Accession: A82197
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-258 <HEI>
 A;Cross-references: GB:AE004224; GB:AE003852; NID:99555952; PIDN:AAF94614.1; GSPDB:GN001
 A;Experimental source: serogroup O1; strain N16961; biotype El Tor
 C;Genetics:
 A;Gene: VCL1457
 A;Map position: 1
 C;Complex: the cholera enterotoxin molecule contains three kinds of chains; an alpha am
 ciate noncovalently with the subunit B, an aggregate of five beta chains
 C;Function:
 A;Description: the active component of the toxin that is primarily responsible for seve
 .2.5) activity also activates intracellular adenylyl cyclase
 C;Superfamily: heat-labile enterotoxin chain A
 C;Keywords: enterotoxin
 F;1-18/Domain: signal sequence #status predicted <SIG>
 F;19-212/Product: cholera enterotoxin alpha chain #status experimental <CTA>
 F;213-258/Product: cholera enterotoxin gamma chain #status experimental <CTG>
 F;217/Diulfide bonds: interchain (to alpha chain) #status predicted

Query Match	43.8%	Score 7;	DB 1;	Length 258;
Best Local Similarity	100.0%	Pred. No. 3.5;		
Matches	7;	Conservative 0;	Mismatches 0;	Indels 0;
Gaps	0;			

Qy 2 SLRSAHL 8
 |||||
 DB 83 SLRSAHL 89

RESULT 3
 S63403
 probable membrane protein YNR070w - yeast (*Saccharomyces cerevisiae*)
 N;Alternate names: hypothetical protein N3568
 C;Species: *Saccharomyces cerevisiae*
 C;Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 19-Apr-2002
 C;Accession: S63403; S63402
 R;Andre, B.; Iraqi Housaini, I.; Urrestazu, L.A.; Vissers, S.
 submitted to the Protein Sequence Database, April 1996
 A;Reference number: S62920
 A;Accession: S63403
 A;Molecule type: DNA
 A;Residues: 1-1333 <AND>
 A;Cross-references: EMBL:Z71685; NID:gl302603; PIDN:CAA96352.1; PID:e239601; PID:gl302606
 A;Experimental source: strain S289C
 R;Duesterhoeft, A.; Floeth, M.; Fritz, C.; Heuss-Neitzel, D.; Hilbert, H.; Moestl, D.
 submitted to the Protein Sequence Database, April 1996
 A;Reference number: S62944
 A;Accession: S63402
 A;Molecule type: DNA
 A;Residues: 1-1053 <DUE>
 A;Cross-references: EMBL:Z71685; MIPS:YNR070w
 A;Experimental source: strain S289C
 C;Genetics:
 A;Cross-references: SGD:S0005353
 A;Map position: 14R
 C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
 C;Keywords: ATP; nucleotide binding; P-loop; transmembrane protein
 F;46-257/Domain: ATP-binding cassette homology <ABC1>
 F;427-443/Domain: transmembrane #status predicted <TM1>
 F;478-494/Domain: transmembrane #status predicted <TM2>
 F;535-551/Domain: transmembrane #status predicted <TM3>
 F;642-658/Domain: transmembrane #status predicted <TM4>
 F;748-946/Domain: ATP-binding cassette homology <ABC2>
 F;765-772/Region: nucleotide-binding motif A (P-loop)

Query Match	43.8%	Score 7;	DB 2;	Length 1333;
-------------	-------	----------	-------	--------------

Query Match 37.5%; Score 6; DB 2; Length 140;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AHLAQ 11
 |||||
 DB 88 AHLAQ 93

RESULT 9
 QOEUC2
 Hypothetical 21.9K protein - Escherichia coli plasmid pCU1
 C:Species: Escherichia coli
 C:Date: 31-Mar-1998 #sequence_revision 30-Jun-1990 #text_change 26-Aug-1999
 C:Accession: JQ0047
 R:Xozloweki, M.; Thatté, V.; Lau, P.C.K.; Visentin, L.P.; Iyer, V.N.
 Gene 58, 217-228, 1997
 A:Title: Isolation and structure of the replicon of the promiscuous plasmid pCU1.
 A:Reference number: JQ0045; MUID:88112872; PMID:2828186
 A:Accession: JQ0047
 A:Molecule type: DNA
 A:Residues: 1-207 <KOZ>
 C:Genetics:
 A:Genome: plasmid
 C:Superfamily: Escherichia coli plasmid pCU1 hypothetical 21.9K protein

Query Match 37.5%; Score 6; DB 1; Length 207;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SAHLAQ 10
 |||||
 DB 28 SAHLAQ 33

RESULT 10
 AG2274
 methyltransferase [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AG2274
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AG2274
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-207 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA075449.1; PID:gl7132884; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all3750
 C:Superfamily: spore germination protein C2; bioC homology

Query Match 37.5%; Score 6; DB 2; Length 207;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AGQSIL 14
 |||||
 DB 189 AGQSIL 194

RESULT 11
 F82729
 conserved hypothetical protein XF1054 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
 C:Accession: F82729
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: F82729
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-232 <SIM>
 A:Cross-references: GB:AE003942; GB:AE003849; NID:99105990; PIDN:AAF83864.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, E
 de Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GeneBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohr
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, F
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.F.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF1054
 C:Superfamily: Streptomyces coelicolor probable integral membrane protein SC6G10.12

Query Match 37.5%; Score 6; DB 2; Length 232;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SAHLAQ 10
 |||||
 DB 206 SAHLAQ 211

RESULT 12
 AF3327
 transcription regulator, tetr family BMEI0604 [imported] - Brucella melitensis (strain 1
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 C:Accession: AF3327
 R:DeLVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujeer, C.; Los, T.; Ivanova,
 Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letese
 proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AF3327
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-241 <KUR>
 A:Cross-references: GB:AE008917; PIDN:AAL51785.1; PID:gl7982528; GSPDB:GN00190
 A:Experimental source: strain 16W
 C:Genetics:
 A:Gene: BMEI0604
 A:Map position: 1

Query Match 37.5%; Score 6; DB 2; Length 241;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LAGQSI 13
 |||||
 DB 203 LAGQSI 208

RESULT 13
 SI2089
 hypothetical protein 245 - Escherichia coli plasmid pCU1
 C:Species: Escherichia coli
 C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Sep-1999
 C:Accession: SI2089

R;Krishnan, B.R.; Fobert, P.R.; Seitzer, U.; Iyer, V.N.
Gene 91, 1-7, 1990
A:Title: Mutations within the replicon of the IncN plasmid pCU1 that affect its Escherichia coli plasmid pCU1 that affect its Escherichia coli plasmid pCU1
A:Reference number: JQ0682; MUID:90382682; PMID:2205534
A:Accession: S12089
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-245 <KRI>
A:Cross-references: EMBL:X52972
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1990
C:Genetics:
A:Genome: plasmid pCU1
C:Superfamily: Escherichia coli plasmid pCU1 hypothetical 21.9K protein

Query Match 37.5%; Score 6; DB 2; Length 245;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SAHLAG 10
Db 28 SAHLAG 33

RESULT 14
C75257
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
C:Accession: C75257
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: C75257
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-270 <WHI>
A:Cross-references: GB:AE002086; GB:AE000513; NID:G6460395; PIDN:AAF12115.1; PID:G646040
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2569
A:Map position: 1
C:Superfamily: Deinococcus radiodurans hypothetical protein DR2569

Query Match 37.5%; Score 6; DB 2; Length 270;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLSRA 6
Db 110 LSLSRA 115

RESULT 15
D90578
elongation factor ts (ef-ts) [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: D90578
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.; Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: D90578
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-294 <KUR>
A:Cross-references: GB:AL445566; PID:G14089947; PIDN:CAC13705.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPU_5320

A:Genetic code: SGC3
C:Superfamily: translation elongation factor EF-Ts

Query Match 37.5%; Score 6; DB 2; Length 294;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 QSILSG 16
Db 225 QSILSG 230

Search completed: April 23, 2004, 14:54:54
Job time : 12.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:47:06 ; Search time 8 Seconds
(without alignments)
104.140 Million cell updates/sec

Title: US-09-528-682-1_COPY_64_79
Perfect score: 16
Sequence: 1 LSLRSALHAGQSILSG 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	16	100.0	258	1 ELAH_ECOLI	P43530 escherichia
2	16	100.0	258	1 ELAP_ECOLI	P06717 escherichia
3	7	43.8	258	1 CHTA_VIBCH	P01555 vibrio chol
4	7	43.8	1333	1 YN99_YEAST	P53756 saccharomyc
5	6	37.5	123	1 RL35_CAEEL	P34662 caenorhabdi
6	6	37.5	245	1 YPC2_ECOLI	P18128 escherichia
7	6	37.5	294	1 EFTS_MYCPU	Q98q37 mycoplasma
8	6	37.5	359	1 HRCB_RHIME	Q926k1 rhizobium m
9	6	37.5	372	1 COLB_ARATH	Q98se5 arabidopsis
10	6	37.5	380	1 METX_THETH	Q9ra51 thetumus the
11	6	37.5	439	1 AROA_DEIRA	Q9rvd3 deinococcus
12	6	37.5	518	1 PTMB_STACA	P28008 staphylococ
13	6	37.5	576	1 YAG3_SCHPO	Q09868 schizosacch
14	6	37.5	617	1 HSCA_VIBVO	Q8dez1 vibrio vuln
15	6	37.5	653	1 CANA_MACFA	Q951p4 macaca fasc
16	6	37.5	663	1 MNE1_YEAST	P24720 saccharomyc
17	6	37.5	696	1 SP15_TORCA	P19965 torpedo cal
18	6	37.5	700	1 GPX1_DROME	P32865 drosophila
19	6	37.5	723	1 S21C_MOUSE	Q8k078 mus musculu
20	6	37.5	723	1 S21C_RAT	Q99n01 rattus norv
21	6	37.5	737	1 YMA2_YEAST	Q04263 saccharomyc
22	6	37.5	830	1 CADG_MOUSE	Q88338 mus musculu
23	6	37.5	908	1 ATMB_SALTY	P22036 salmonella
24	6	37.5	1501	1 SNQ2_YEAST	P32568 saccharomyc
25	5	31.2	47	1 R362_YERPE	Q8zc86 yersinia pe
26	5	31.2	56	1 RL32_SYNY3	P73014 synchocyst
27	5	31.2	70	1 IF1_MYCGE	P47419 mycoplasma
28	5	31.2	76	1 UCRX_YEAST	P37299 saccharomyc
29	5	31.2	78	1 IF1_MYCPN	Q50298 mycoplasma
30	5	31.2	86	1 RL27_BORBR	Q7wq17 bordetella
31	5	31.2	86	1 RL27_BORPA	Q7wip1 bordetella
32	5	31.2	86	1 RL27_BORPE	Q7vzx5 bordetella
33	5	31.2	86	1 VGC_BPPHX	P03635 bacteriopha

RESULT 1

ELAH_ECOLI 94 1 RL23 TREPA 083221 treponema p
35 5 31.2 96 1 CH10_CANTP Q8ktr9 candidatus
36 5 31.2 99 1 YIIS_ECOLI P32162 escherichia
37 5 31.2 99 1 YQJK_ECOLI Q47710 escherichia
38 5 31.2 108 1 Y267_VIBMA Q9ra13 vibrio mari
39 5 31.2 115 1 DAD1_ARATH Q39080 arabidopsis
40 5 31.2 115 1 DAD1_BETVE Q9m3t9 betula verr
41 5 31.2 115 1 DAD1_CITUN Q9zwg7 citrus unsh
42 5 31.2 115 1 DAD1_PICMA Q65085 picea maria
43 5 31.2 116 1 DAD1_LYCES Q9smc4 lycopersico
44 5 31.2 118 1 Y759_METJA Q58169 methanococc
45 5 31.2 119 1 DAD1_MALDO O24060 malus domes

ALIGNMENTS

ID ELAH_ECOLI STANDARD; PRT; 258 AA.
AC P43530;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heat-labile enterotoxin A chain precursor (LT-A, human) (LTH-A).
GN ELTA OR LTPA OR TOXA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID:562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O78:H11 / H10407;
RX MEDLINE=84185610; PubMed=6325417;
RA Yamamoto T., Tamura T., Yokota T.;
RT "Primary structure of heat-labile enterotoxin produced by Escherichia coli pathogenic for humans.";
RL J. Biol. Chem. 259:5037-5044(1984).
RN [2]
RP REVISION TO 207.
RC STRAIN=O78:H11 / H10407;
RX MEDLINE=87137303; PubMed=3546273;
RA Yamamoto T., Gojobori T., Yokota T.;
RT "Evolutionary origin of pathogenic determinants in enterotoxigenic Escherichia coli and Vibrio cholerae O1.";
RJ J. Bacteriol. 169:1352-1357(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O78:H11 / H10407;
RX MEDLINE=93252225; PubMed=8486242;
RA Inoue T., Tsuji T., Koto M., Imamura S., Miyama A.;
RT "Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigenic Escherichia coli is identical to that of human strain H 10407.";
RJ FEMS Microbiol. Lett. 108:157-161(1993).
RN [4]
RP DISCUSSION OF SEQUENCE.
RA Domenighini M., Pizza M., Jobling M.G., Holmes R.K., Rappuoli R.;
RT "Identification of errors among database sequence entries and comparison of correct amino acid sequences for the heat-labile enterotoxins of Escherichia coli and Vibrio cholerae.";
RJ Mol. Microbiol. 15:1165-1167(1995).
CC -!- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC -!- SUBUNIT: HETEROHXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)


```

FT TURN 43 44
FT TURN 48 49
FT TURN 51 52
FT HELIX 59 64
FT STRAND 67 67
FT TURN 68 69
FT STRAND 70 70
FT STRAND 77 81
FT HELIX 84 94
FT TURN 96 97
FT STRAND 100 106
FT TURN 110 111
FT STRAND 112 114
FT HELIX 115 119
FT HELIX 120 122
FT TURN 127 128
FT STRAND 130 134
FT TURN 135 135
FT STRAND 137 138
FT HELIX 139 141
FT STRAND 142 149
FT TURN 150 151
FT STRAND 152 159
FT TURN 161 162
FT HELIX 165 168
FT TURN 169 170
FT STRAND 174 174
FT HELIX 176 178
FT TURN 179 179
FT HELIX 180 182
FT TURN 187 188
FT HELIX 190 193
FT TURN 195 196
FT HELIX 197 200
FT TURN 203 204
FT HELIX 215 240
FT HELIX 241 244
FT HELIX 250 253
SQ SEQUENCE 258 AA; 23902 MW; 2F0786442619F81F CRC64;

Query Match 100.0%; Score 16; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 7.8e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LSRSAHLAGQSILSG 16
DB 82 LSRSAHLAGQSILSG 97

RESULT 3
ID CHTA_VIBCH STANDARD; PRT; 258 AA.
AC P01555; Q56634; Q90PVI;
DT 21-JUL-1986 (Rel. 01, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cholera enterotoxin, A chain precursor (NAD(+)-diphthamide ADP-
DE ribosyltransferase) (EC 2.4.2.36) (Cholera enterotoxin A subunit).
GN CTXA OR TOXA OR VC1457.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=El Tor 2125;
RX MEDLINE=84068199; PubMed=6646234;
RA Mekalanos J.J., Swartz D.J., Pearson G.D.N., Harford N., Groyne F.,
RA de Wilde M.;
RT "Cholera toxin genes: nucleotide sequence, deletion analysis and
RL vaccine development.";
RL Nature 306:551-557(1983).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=Classical 569B / ATCC 25870 / Serotype O1;
RX MEDLINE=91355224; PubMed=1883840;
RA Dams E., de Wolf M., Dierick W.;
RT "Nucleotide sequence analysis of the CT operon of the Vibrio cholerae
RL classical strain 569B.";
RN Biochim. Biophys. Acta 1090:139-141(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1854 / O139-Bengal;
RA Yamamoto K., Do V.G.R.F., Xu M., Iida T., Miwatani T., Albert M.J.,
RA Honda T.;
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor 2125;
RA Dams E., de Wolf M., Dierick W.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=KNIH002;
RA Shin H.J., Park Y.C., Kim Y.C.;
RT "Cloning and nucleotide sequence analysis of the virulence gene
RL cassette from Vibrio cholerae KNIH002 isolated in Korea.";
RN Misainmurhag Hoiji 35:205-210(1999).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Wenter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RL cholerae.";
RN Nature 406:477-483(2000).
RN [7]
RP SEQUENCE OF 1-212 FROM N.A.
RC STRAIN=Classical 569B / ATCC 25870 / Serotype O1;
RX MEDLINE=85006737; PubMed=6090390;
RA Lockman H.A., Galen J.E., Kaper J.B.;
RT "Vibrio cholerae enterotoxin genes: nucleotide sequence analysis of
RL DNA encoding ADP-ribosyltransferase.";
RN J. Bacteriol. 159:1086-1089(1984).
RN [8]
RP SEQUENCE OF 213-258 FROM N.A.
RX MEDLINE=84061784; PubMed=6315707;
RA Lockman H., Kaper J.B.;
RT "Nucleotide sequence analysis of the A2 and B subunits of Vibrio
RL cholerae enterotoxin.";
RN J. Biol. Chem. 258:13722-13726(1983).
RN [9]
RP SEQUENCE OF 19-27.
RX MEDLINE=81212799; PubMed=7238869;
RA Duffy L.K., Peterson J.W., Kurosky A.;
RT "Isolation and characterization of a precursor form of the 'A'
RL subunit of cholera toxin.";
RN FEBS Lett. 126:187-190(1981).
RN [10]
RP SEQUENCE OF 19-38 AND 213-232.
RX MEDLINE=76259136; PubMed=955672;
RA Klapper D.G., Finkelstein R.A., Capra J.D.;
RT "Subunit structure and N-terminal amino acid sequence of the three
RL chains of cholera enterotoxin.";
RN Immunochimistry 13:605-611(1976).
RN [11]
RP SEQUENCE OF 27-72 AND 111-139.
RX MEDLINE=79169830; PubMed=437113;
RA Lai C.-Y., Cancedda F., Chang D.;
RT "Primary structure of cholera toxin subunit A1: isolation, partial
RN

```

```

sequences and alignment of the BrCN fragments." ;
RL FBS Lett. 100:85-89(1979).
RN [12]
RP SEQUENCE OF 213-258.
RX MEDLINE=82053094; PubMed=7028752;
RA Duffy L.K., Peterson J.W., Kurosky A.;
RT "Covalent structure of the gamma chain of the A subunit of cholera
toxin." ;
RL J. Biol. Chem. 256:12252-12256(1981).
RN [13]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE=95387395; PubMed=7658473;
RA Zhang R.G., Scott D.L., Westbrook M.L., Nance S., Spangler B.D.,
RT Shipley G.G., Westbrook E.M.;
RT "The three-dimensional crystal structure of cholera toxin." ;
RL J. Mol. Biol. 251:563-573(1995).
RN [14]
RP FUNCTION: THE ALPHA/GAMMA CHAIN (A SUBUNIT) IS AN ADP-RIBOSYLATING
TOXIN.
CC -|- CATALYTIC ACTIVITY: NAD(+) + peptide diphthamide = nicotinamide +
peptide N-(ADP-D-riboseyl)diphthamide.
CC -|- SUBUNIT: CONTAINS 3 KINDS OF CHAINS, AN ALPHA AND A GAMMA CHAIN
(PROM THE SAME PRECURSOR MOLECULE), LINKED BY AN INTERCHAIN
DISULFIDE BOND, ASSOCIATE NONCOVALENTLY WITH AN AGGREGATE OF 4 TO
6 BETA CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X00171; CAA24995.1; -
DR EMBL; X58785; CAA41590.1; -
DR EMBL; D30053; BAA06280.1; -
DR EMBL; X58786; CAA41592.1; -
DR EMBL; AF175708; AAD51359.1; -
DR EMBL; AE004224; AAF94614.1; -
DR EMBL; K01170; AAA27572.1; -
DR EMBL; D30052; BAA06288.1; -
DR PIR; A05129; XVCVA.
DR PDB; IXPB; 01-APR-97.
DR PDB; IXTC; 01-AUG-96.
DR TIGR; VC1457; -
DR InterPro; IPR001144; Enterotoxin_A.
DR Pfam; PF01375; Enterotoxin_A; 1.
DR PRINTS; PR00771; ENTEROTOXINA.
KW Enterotoxin; Signal; NAD; Transferase; Glycosyltransferase;
KW 3D-structure; Complete proteome.
FT SIGNAL 1 18
FT CHAIN 19 212 CHOLERA ENTEROTOXIN, CHAIN-A1 (ALPHA).
FT CHAIN 213 258 CHOLERA ENTEROTOXIN, CHAIN-A2 (GAMMA).
FT DISULFD 217 217 INTERCHAIN (WITH GAMMA CHAIN).
FT ACT_SITE 62 62 INTERACT WITH NAD (BY SIMILARITY).
FT ACT_SITE 130 130 BY SIMILARITY.
FT CONFLICT 20 20 D -> N (IN REF. 9).
FT CONFLICT 37 37 S -> R (IN REF. 10).
FT CONFLICT 39 39 G -> L (IN REF. 11).
FT CONFLICT 45 46 QS -> SE (IN REF. 11).
FT CONFLICT 111 111 N -> L (IN REF. 11).
FT CONFLICT 132 132 S -> A (IN REF. 11).
FT CONFLICT 213 213 M -> I (IN REF. 11).
FT CONFLICT 247 248 DI -> ID (IN REF. 12).
FT CONFLICT 256 256 D -> N (IN REF. 12).
FT STRAND 24 27
FT HELIX 31 37
FT TURN 38 38
FT STRAND 39 40
FT TURN 43 44
FT TURN 48 49
FT HELIX 59 63

```

```

FT TURN 64 64
FT TURN 75 76
FT STRAND 77 81
FT HELIX 85 89
FT TURN 90 91
FT TURN 96 97
FT STRAND 101 106
FT TURN 110 111
FT STRAND 112 114
FT HELIX 115 119
FT HELIX 120 122
FT HELIX 126 128
FT STRAND 130 134
FT STRAND 137 138
FT TURN 139 141
FT STRAND 142 148
FT STRAND 153 159
FT TURN 161 162
FT HELIX 165 168
FT TURN 169 170
FT HELIX 176 178
FT TURN 187 188
FT HELIX 190 193
FT TURN 195 196
FT HELIX 197 199
FT TURN 200 200
FT TURN 203 204
FT HELIX 215 251
FT TURN 252 253
FT HELIX 254 258
SQ SEQUENCE 258 AA; 29335 MW; 0F7EBAE62069A5D0 CRC64;

Query Match 43.8%; Score 7; DB 1; Length 258;
Best Local Similarity 100.0%; Pred.No. 2.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 SLRSAHL 8
Db 83 SLRSAHL 89

RESULT 4
YN99 YEAST STANDARD; PRT; 1333 AA.
AC P53756;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Probable ATP-dependent transporter YNR070W.
GN YNR070W OR N3568.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
RN [1] NCBI_TaxID=4932;
RP SEQUENCE FROM N.A.
RA Andre B., Iraqi Houssaini I., Urrestarazu L.A., Vissers S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-1054 FROM N.A.
RA Dueterhoeft A., Floeth M., Fritz C., Heuss-Neitzel D.,
RA Hilbert H., Moestl D.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -|- SIMILARITY: Belongs to the ABC transporter family. PDR5 subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----

```

```

CC -----
DR EMBL; Z71685; CAA96352.1; -.
DR EMBL; Z71686; CAA96354.1; -.
DR FIR; S63403; S63403.
DR GERMOnline; 143415; -.
DR SGD; S0005353; YNR070W.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transporter; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
KW Hypothetical protein; ATP-binding; Transmembrane; Glycoprotein;
KW Transport.
FT TRANSMEM 13 33 POTENTIAL.
FT TRANSMEM 392 412 POTENTIAL.
FT TRANSMEM 425 445 POTENTIAL.
FT TRANSMEM 474 494 POTENTIAL.
FT TRANSMEM 499 519 POTENTIAL.
FT TRANSMEM 534 554 POTENTIAL.
FT TRANSMEM 642 662 POTENTIAL.
FT TRANSMEM 1071 1091 POTENTIAL.
FT TRANSMEM 1092 1112 POTENTIAL.
FT TRANSMEM 1150 1170 POTENTIAL.
FT TRANSMEM 1178 1198 POTENTIAL.
FT TRANSMEM 1210 1230 POTENTIAL.
FT TRANSMEM 1235 1255 POTENTIAL.
FT NP_BIND 765 772 ATP (POTENTIAL).
FT CARBOHYD 48 48 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 205 205 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 350 350 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 697 697 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 733 733 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 958 958 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1320 1320 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1333 AA; 149749 MW; 61B4758E0245CB70 CRC64;

Query Match 43.8%; Score 7; DB 1; Length 1333;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LAGQSIL 14
DB 917 LAGQSIL 923
|||||

RESULT 5
RL35 CAEL STANDARD; PRT; 123 AA.
AC P34652;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L35.
GN RPL-35 OR ZK652.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonnenhammer E., Staden R.,
RA S.

```

```

RA Sulston J., Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -|- SIMILARITY: Belongs to the L29P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L14429; AAA28216.1; -.
CC FIR; S44905; S44905.
CC WormPep; ZK652.4; CE00450.
CC InterPro; IPR001854; Ribosomal L29.
CC Pfam; PF00831; Ribosomal L29; 1.
CC TIGRFAMs; TIGR00012; L29; 1.
CC PROSITE; PS00579; RIBOSOMAL_L29; 1.
CC KW Ribosomal protein.
SQ SEQUENCE 123 AA; 14195 MW; 30F800D03DB05F27 CRC64;

Query Match 37.5%; Score 6; DB 1; Length 123;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLSRSA 6
DB 100 LLSRSA 105
|||||

RESULT 6
YPC2 ECOLI STANDARD; PRT; 245 AA.
ID YPC2_ECOLI
AC P18128;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Hypothetical 26.4 kDa protein (ORF 245).
OS Escherichia coli.
OG Plasmid IncN pCU1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KL2;
RX MEDLINE=90382682; PubMed=2205534;
RA Krishnan B.R., Fobert P.R., Seitzer U., Iyer V.N.;
RA "Mutations within the replicon of the IncN plasmid pCU1 that affect
RA its Escherichia coli polA-independence but not its autonomous
RA replication ability.";
RL Gene 91:1-7(1990).
RN [2]
RP SEQUENCE OF 1-207 FROM N.A.
RX MEDLINE=88112872; PubMed=288186;
RA Kozlowski M., Thatte V., Lau P.C.K., Visentin L.P., Iyer V.N.;
RA "Isolation and structure of the replicon of the promiscuous plasmid
RA pCU1.";
RL Gene 58:217-228(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; M18262; AAA98072.1; --
DR EMBL; M18262; AAA98071.1; ALT_INIT.
DR PIR; JQ0047; QOECU2.
KW Hypothetical protein; Plasmid.
FT CONFLICT 206 207 AS -> GK (IN REF. 2).
SQ SEQUENCE 245 AA; 26287 MW; 550809B6577FBB5 CRC64;

Query Match 37.5%; Score 6; DB 1; Length 245;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SAHLAG 10
    |||||
    28 SAHLAG 33

RESULT 7
EFTS MYCPU STANDARD; PRT; 294 AA.
ID EFTS MYCPU STANDARD; PRT; 294 AA.
AC Q98Q37;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Elongation factor Ts (EF-Ts).
GN TSF OR MYPU 5320.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN PROTEIN: Negative regulator of class I heat shock genes [grp8-
  dnak-dna and groELS operons]. Prevents heat-shock induction of
  these operons (By similarity).
  -1- SIMILARITY: Belongs to the EF-Ts family.
  This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
  -1- SUBCELLULAR LOCATION: Cytoplasmic.
  -1- SIMILARITY: Belongs to the EF-Ts family.
  EMBL; AL445565; CAC13705.1; --
  PIR; D90578; D90578.
  MYPULIST; MYPU_5320; --
  HAMAP; MF_00050; --; 1.
  InterPro; IPR001816; EF_TS.
  DR InterPro; IPR000449; UBA_domain.
  DR Pfam; PF00889; EF_TS; 1.
  DR Pfam; PF00627; UBA; 1.
  DR TIGRFAMs; TIGR00116; tsf; 1.
  DR PROSITE; PS01126; EF_TS_1; 1.
  DR PROSITE; PS01127; EF_TS_2; 1.
  KW Elongation factor; Protein biosynthesis; Complete proteome.
  FT SITE 81 84 INVOLVED IN MG++ ION DISLOCATION FROM EF-
  TU (BY SIMILARITY).
  FT SEQUENCE 294 AA; 32627 MW; 192CCCECF499A41 CRC64;
  Query Match 37.5%; Score 6; DB 1; Length 294;
  Best Local Similarity 100.0%; Pred. No. 29;
  Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 QSILSG 16

```

```

Db 225 QSILSG 230
    |||||
    225 QSILSG 230

RESULT 8
HRCA RHIME STANDARD; PRT; 359 AA.
ID HRCA RHIME STANDARD; PRT; 359 AA.
AC Q92SK1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heat-inducible transcription repressor hrca.
GN HRCA OR R00377 OR SMC01143.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN PROTEIN: Negative regulator of class I heat shock genes [grp8-
  dnak-dna and groELS operons]. Prevents heat-shock induction of
  these operons (By similarity).
  -1- SIMILARITY: Belongs to the hrca family.
  This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
  EMBL; AL591783; CAC41814.1; --
  HAMAP; MF_00081; --; 1.
  InterPro; IPR002571; Hrca.
  DR Pfam; PF01628; Hrca; 1.
  DR TIGRFAMs; TIGR00331; hrca; 1.
  KW Transcription regulation; Repressor; Heat shock; Complete proteome.
  SQ SEQUENCE 359 AA; 39168 MW; C02450A151999CD CRC64;
  Query Match 37.5%; Score 6; DB 1; Length 359;
  Best Local Similarity 100.0%; Pred. No. 35;
  Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 AHLAQ 11
    |||||
    197 AHLAQ 202

RESULT 9
COLB ARATH STANDARD; PRT; 372 AA.
ID COLB ARATH STANDARD; PRT; 372 AA.
AC Q9SSE5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Zinc finger protein constans-like 11.
GN AT3G07650 OR MLP3.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

```


RN SEQUENCE FROM N.A.
 RC STRAIN=CV, Columbia;
 EX MEDLINE=21016720; PubMed=11130713;
 RA Salanoubat M., Iemcke K., Rieger M., Anserge W., Unselid M.,
 RA Farmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA Delisny M., Boutry M., Grievell L.A., Mache R., Puigdomenech P.,
 RA De Simone V., Choiane N., Artiguenave F., Robert C., Brottier P.,
 RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurnbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Verzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
 RA Reichelt J., Schaefer M., Schoen M., Schone M., Terol J., Climent J.,
 RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
 RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
 RA Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,
 RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
 RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 RA Pal G., Miltscher J., Sellers P., Gill J.E., Feldblyum T.V.,
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakayama S., Nakazaki N., Shimpō S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 thaliana";
 RL Nature 408:820-822 (2000).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- SIMILARITY: Belongs to the CONSTANS family.
 CC -!- SIMILARITY: Contains 2 B box-type zinc fingers.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AC009176; AAF13083.1; -
 DR InterPro; IPR000315; Znf.Bbox.
 DR InterPro; IPR002926; Znf.CONSTANS.
 DR Pfam; PF00643; ZF-B_box; 2.
 DR ProDom; PD007661; Znf.CONSTANS; 1.
 DR SMART; SM00336; BBOX; 2.
 DR PROSITE; PS01119; ZF_BBOX; 2.
 KW Zinc-finger; Nuclear protein; Repeat; Multigene family.
 FT ZN_FING 5 47 B_BOX-TYPE 1.
 FT ZN_FING 48 99 B_BOX-TYPE 2 (ATYPICAL).
 FT DOMAIN 77 83 POLY-ASN.
 FT DOMAIN 84 90 POLY-SER.
 FT DOMAIN 90 90 POLY-SER.
 SQ SEQUENCE 372 AA; 40754 MW; 188F18EB283D7479 CRC64;
 Query Match 37.5%; Score 6; DB 1; Length 372;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 LAGQSI 13
 DB 121 LAGQSI 126
 RESULT 10
 METX THETH
 ID METX THETH STANDARD; PRT; 380 AA.
 AC Q9RA51.
 DT 16-OCT-2001 (Rel. 40, Created)

DT SEQUENCE FROM N.A.
 DT STRAIN=CV, Columbia;
 DE MEDLINE=21016720; PubMed=11130713;
 GN Salanoubat M., Iemcke K., Rieger M., Anserge W., Unselid M.,
 OS Farmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 OC Delisny M., Boutry M., Grievell L.A., Mache R., Puigdomenech P.,
 OC De Simone V., Choiane N., Artiguenave F., Robert C., Brottier P.,
 RX Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
 RX Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RP Wurnbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
 RC Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Verzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
 RT Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
 RT Reichelt J., Schaefer M., Schoen M., Schone M., Terol J., Climent J.,
 RL Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
 CC Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
 CC de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
 CC Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,
 CC Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
 CC Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 CC Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
 CC Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 CC Pal G., Miltscher J., Sellers P., Gill J.E., Feldblyum T.V.,
 CC Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
 CC Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 CC Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
 CC Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 CC Nakayama S., Nakazaki N., Shimpō S., Takeuchi C., Wada T.,
 CC Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 thaliana";
 RL Nature 408:820-822 (2000).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- SIMILARITY: Belongs to the CONSTANS family.
 CC -!- SIMILARITY: Contains 2 B box-type zinc fingers.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB029372; BAA88676.1; -
 DR HAMAP; MF 00296; - 1.
 DR InterPro; IPR000073; A/b hydrolase.
 DR InterPro; IPR003089; AB_Hydrolase.
 DR InterPro; IPR008220; Homoser_Ac_trans.
 DR InterPro; IPR006296; HomoserO_Ac_trn.
 DR InterPro; IPR000379; Ser_estra.
 DR Pfam; PF00561; abhydrolase; 1.
 DR PIRSF; PIRSF00443; Homoser_Ac_trans; 1.
 DR PRINTS; PF00111; ABHYDROLASE.
 DR TIGRPFAMs; TIGR01392; homoserO_Ac_trn; 1.
 KW Methionine biosynthesis; Transferase; Acyltransferase.
 FT ACT_SITE 186 186 POTENTIAL.
 FT ACT_SITE 361 361 POTENTIAL.
 SQ SEQUENCE 380 AA; 42250 MW; DCFE133CDE933C23 CRC64;
 Query Match 37.5%; Score 6; DB 1; Length 380;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 SAHLAG 10
 DB 81 SAHLAG 86
 RESULT 11
 AROA DEIRA
 ID AROA DEIRA STANDARD; PRT; 439 AA.
 AC Q9RVD3;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
 DR enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
 GN AROA OR DR1096.
 OS Deinococcus radiodurans.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
 RX MEDLINE=20036896; PubMed=10567266;

```

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioreistant bacterium Deinococcus
RT radiodurans RL";
RT Science 286:1571-1577(1999).
CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: Belongs to the EPSP synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001959; AAF10666.1; ALT_INIT.
DR TIGR; DR1036; -.
DR HAMAP; MF_00210; -.
DR InterPro; IPR006264; AroA.
DR InterPro; IPR001986; EPSP synth.
DR Pfam; PF00275; EPSP synthase; 1.
DR ProDom; PD001867; EPSP_synthase; 1.
DR TIGRFAMs; TIGR01356; aroA; 1.
DR PROSITE; PS00104; EPSP SYNTHASE 1; 1.
DR PROSITE; PS00885; EPSP SYNTHASE 2; FALSE NEG.
KW Aromatic amino acid biosynthesis; Transferase; Complete proteome.
SQ SEQUENCE 439 AA; 46673 MW; B30D808D247C5A8B CRC64;

Query Match 37.5%; Score 6; DB 1; Length 439;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SAHLAG 10
DB 374 SAHLAG 379
|||||

RESULT 12
PTMB STACA
ID PTMB STACA STANDARD; PRT; 518 AA.
AC P28008;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE PTS system, mannitol-specific IIBC component (EIBC-Mtl) (Mannitol-
DE permease IIBC component) (Phosphotransferase enzyme II, BC component)
DE (EC 2.7.1.69) (EII-Mtl).
DN MTLA.
OS Staphylococcus carnosus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1281;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92201209; PubMed=1551396;
RA Fischer R., Hengstenberg W.;
RT "Mannitol-specific enzyme II of the phosphoenolpyruvate-dependent
RT phosphotransferase system of Staphylococcus carnosus. Sequence and
RT expression in Escherichia coli and structural comparison with the
RT enzyme IImannitol of Escherichia coli.";
RL Eur. J. Biochem. 204:963-969(1992).
RN [2]
RP REVISIONS.

RA Hengstenberg W.;
RA Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: This is a component of the phosphoenolpyruvate-dependent
CC sugar phosphotransferase system (PTS), a major carbohydrate active
CC -transport system. The IIC domains contain the sugar binding site
CC and the transmembrane channel; the IIA domain contains the primary
CC phosphorylation site (the donor is phospho-HPr); IIA transfers its
CC phosphoryl group to the IIB domain which finally transfers it to
CC the sugar.
CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC histidine + sugar phosphate.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- INDUCTION: BY MANNITOL.
CC -!- SIMILARITY: Contains 1 PTS EIIA domain.
CC -!- SIMILARITY: Contains 1 PTS EIIB domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X56333; CAA39769.1; -.
DR PIR; S68193; S22385.
DR InterPro; IPR003352; Ptrans_EIIC.
DR InterPro; IPR003501; Ptrans_IIB.
DR InterPro; IPR004718; PTS_EIIC_mtlA.
DR Pfam; PF02378; PTS_EIIC; 1.
DR Pfam; PF02302; PTS_IIB; 1.
DR TIGRFAMs; TIGR00851; mtlA; 1.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Transmembrane; Phosphorylation.
FT DOMAIN 1 342
FT DOMAIN 343 518
FT EIIB.
FT DOMAIN 1 32
FT CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 33 52
FT POTENTIAL.
FT DOMAIN 53 58
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 59 77
FT POTENTIAL.
FT DOMAIN 78 143
FT CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 144 163
FT POTENTIAL.
FT DOMAIN 164 174
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 175 193
FT POTENTIAL.
FT DOMAIN 194 282
FT CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 283 300
FT POTENTIAL.
FT DOMAIN 301 322
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 323 342
FT POTENTIAL.
FT DOMAIN 343 518
FT CYTOPLASMIC (PROBABLE).
FT MOD_RES 265 265
FT PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 432 432
FT PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 518 AA; 55685 MW; A0DE09E4BA74FA5 CRC64;

Query Match 37.5%; Score 6; DB 1; Length 518;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AQOSIL 14
DB 218 AQOSIL 223
|||||

RESULT 13
YAG3 SCHPO
ID YAG3 SCHPO STANDARD; PRT; 576 AA.
AC Q09868;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical RNA-binding protein Cl2G12.03 in chromosome I.
DE SPAC12G12.03.
OS Schizosaccharomyces pombe (Fission yeast).
RP REVISIONS.

```

```

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=2184801; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sproules J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moulis S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann M., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RA "The genome sequence of Schizosaccharomycetes pombe.";
RL Nature 415:871-880(2002).
CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC -!- SIMILARITY: SOME, TO YEAST YBL051C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z66568; CAA91498.1; -.
DR FIR; S62534; S62534.
DR GenBank; SPombe; SPAC12G12.03; -.
DR InterPro; IPR001374; R3H.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01424; R3H; 1.
DR Pfam; PF00076; xrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS00102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Hypothetical protein; RNA-binding.
PT DOMAIN 232 310 RNA-BINDING (RRM).
SQ SEQUENCE 576 AA; 62068 MW; 41FBD27201EE7D07 CRC64;
Query Match 37.5%; Score 6; DB 1; Length 576;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LSLRSA 6
DB 157 LSLRSA 162
RESULT 14
HSCA_VIBVU
ID HSCA_VIBVU STANDARD; PRT; 617 AA.
AC Q8DE21;
DT 10-OCT-2003 (Rel. 42, Created)

```

```

DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Chaperone protein hscA homolog.
GN HSCA OR VJ10434.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.B.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Probable chaperone. Has a low intrinsic ATPase activity
CC which is markedly stimulated by hscB (By similarity).
CC -!- SIMILARITY: Belongs to the heat shock protein 70 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB016798; AB008957.1; -.
DR HAMAP; MF_00679; -.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR ProDom; PD000089; Hsp70; 1.
DR PROSITE; PS00297; HSP70; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; FALSE NEG.
KW Chaperone; ATP-binding; Complete proteome.
SQ SEQUENCE 617 AA; 66094 MW; 431A85D10B09BE87 CRC64;
Query Match 37.5%; Score 6; DB 1; Length 617;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 LAGQSI 13
DB 609 LAGQSI 614
RESULT 15
CANA_MACFA
ID CANA_MACFA STANDARD; PRT; 653 AA.
AC Q95LP4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Calpain 10 (EC 3.4.22.-) (Calcium-activated neutral proteinase 10)
DE (CAMP 10).
GN CAPN10.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Testis;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.;
RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT libraries.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in

```

```

CC cytoskeletal remodelling and signal transduction.
CC -1- CATALYTIC ACTIVITY: Broad endopeptidase specificity.
CC -1- SIMILARITY: Belongs to peptidase family C2.
CC -1- SIMILARITY: Contains 1 Calpain catalytic domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB072744; BAB69713.1; -.
CC InterPro; IPR001300; Peptidase_C2.
CC InterPro; IPR000169; SHprot_acsite.
CC Pfam; PF01067; Calpain_III; 2.
CC Pfam; PF00648; Peptidase_C2; 1.
CC PRINTS; PR00704; CALPAIN.
CC SMART; SM00720; calpain_III; 2.
CC SMART; SM00230; Cyspc; 1.
CC PROSITE; PS50203; CALPAIN_CAT; 1.
CC PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
CC PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
CC PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
CC KX Hydrolase; Thiol protease; Repeat.
CC FT DOMAIN 13 321 CALPAIN CATALYTIC.
CC FT DOMAIN 322 494 DOMAIN_III_1.
CC FT DOMAIN 513 653 DOMAIN_III_2.
CC FT ACT_SITE 73 73 BY SIMILARITY.
CC FT ACT_SITE 238 238 BY SIMILARITY.
CC FT ACT_SITE 263 263 BY SIMILARITY.
CC SQ SEQUENCE 653 AA; 72997 MW; 47812497BB315971 CRC64;

Query Match 37.5%; Score 6; DB 1; Length 653;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 AGQSIL 14
Db 252 AGQSIL 257
|||||

```

Search completed: April 23, 2004, 14:52:13
Job time : 9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:47:36 ; Search time 33.3333 Seconds
(without alignments)

151.449 Million cell updates/sec

Title: US-09-528-682-1_COPY_64_79

Perfect score: 16

Sequence: 1 LSLRSAHLAQSLSG 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	100.0	258	2 O66280	O66280 escherichia
2	7	43.8	194	16 Q87KL8	Q87KL8 vibrio para
3	7	43.8	194	16 Q7V4Z6	Q7V4Z6 prochloroco
4	7	43.8	258	2 Q8L356	Q8L356 vibrio chol
5	7	43.8	258	2 Q8VLJ6	Q8VLJ6 vibrio chol
6	7	43.8	258	9 Q8L1G8	Q8L1G8 vibrio phag
7	7	43.8	310	16 Q7UNS6	Q7UNS6 rhodospirell
8	7	43.8	335	2 Q8RPJ7	Q8RPJ7 desulfitoba
9	7	43.8	360	16 Q7UHC6	Q7UHC6 rhodospirell
10	7	43.8	561	5 Q9VIM8	Q9VIM8 drosophila
11	7	43.8	577	16 Q8J8J9	Q8J8J9 enterococcu
12	7	43.8	580	5 Q8SZ04	Q8SZ04 drosophila
13	7	43.8	580	5 Q917S3	Q917S3 drosophila
14	7	43.8	842	16 Q88DQ1	Q88DQ1 pseudomonas
15	7	43.8	1040	16 Q82WW7	Q82WW7 nitrosomona
16	7	43.8	1065	4 O94898	O94898 homo sapien

17	6	37.5	72	2	Q93JN6
18	6	37.5	73	16	Q7V5C5
19	6	37.5	86	8	Q35321
20	6	37.5	91	12	Q8QNP9
21	6	37.5	93	16	Q7VKZ3
22	6	37.5	110	16	Q8NLC8
23	6	37.5	114	2	P72225
24	6	37.5	116	10	Q9AVB2
25	6	37.5	120	16	Q912Q5
26	6	37.5	128	17	Q9YEQ0
27	6	37.5	133	4	Q9HAA7
28	6	37.5	142	2	Q83XV2
29	6	37.5	145	4	Q96MQ1
30	6	37.5	157	16	Q82X36
31	6	37.5	194	2	Q9AHP3
32	6	37.5	202	16	Q8PAS1
33	6	37.5	207	16	Q8YQR9
34	6	37.5	220	5	Q8IP84
35	6	37.5	224	16	Q8PMH7
36	6	37.5	232	16	Q9PEH4
37	6	37.5	232	16	Q87EH4
38	6	37.5	241	16	Q8Y141
39	6	37.5	241	16	Q8FZS0
40	6	37.5	241	16	Q7VIT5
41	6	37.5	245	16	Q7UMC3
42	6	37.5	247	10	Q8L8D8
43	6	37.5	247	16	Q82UE8
44	6	37.5	261	16	Q8XLY9
45	6	37.5	269	2	Q9L528

ALIGNMENTS

RESULT 1

AC	O66280	PRELIMINARY;	PRT;	258 AA.	
ID	O66280				
OC	O66280;				
DT	01-AUG-1998	(TrEMBLrel. 07, Created)			
DT	01-AUG-1998	(TrEMBLrel. 07, Last sequence update)			
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)			
DE	Heat-labile enterotoxin A subunit.				
GN	LTH A SUBUNIT.				
OS	Escherichia coli.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Escherichia.				
OX	NCBI_TaxID=562;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=1032;				
RA	Komase K.;				
RL	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=1032;				
RX	MEDLINE=95091056; PubMed=7998417;				
RA	Tamura S., Asanuma H., Tomita T., Komase K., Kawahara K., Danbara				
RA	Hattori N., Watanabe K., Suzuki Y., Nagamine T., Aizawa C., Oya A.,				
RA	Kurata T.;				
RT	"Escherichia coli heat-labile enterotoxin B subunits supplemented				
RT	a trace amount of the holotoxin as an adjuvant for nasal influenza				
RT	vaccine.";				
RL	Vaccine 12:1083-1089(1994).				
DR	EMBL; AB011677; BAA25725.1; -.				
DR	HSSP; P06717; 1LTG.				
DR	GO; GO:0005576; C:extracellular; IEA.				
DR	GO; GO:0015070; F:toxin activity; IEA.				
DR	GO; GO:0009405; P:pathogenesis; IEA.				
DR	InterPro; IPR001144; Enterotoxin_A.				
DR	InterPro; IPR000866; ER_target_S.				
DR	Pfam; PF01375; Enterotoxin_A; 1.				
DR	PRINTS; PR00771; ENTEROTOXINA.				
DR	PROSITE; PS00014; ER_TARGET; 1.				

SQ SEQUENCE 258 AA; 29931 MW; 2BB15D27740EB788 CRC64;

Query Match 100.0%; Score 16; DB 2; Length 258;
Best Local Similarity 100.0%; Pred. No. 6.6e-09;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLSRAHLAQSSILSG 16

|||||

Db 82 LSLSRAHLAQSSILSG 97

RESULT 2

Q87KL8 PRELIMINARY; PRT; 194 AA.

AC Q87KL8;

DT 01-JUN-2003 (TREMBlrel. 24, Created)

DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Putative membrane protein.

GN VP2959.

OS Vibrio parahaemolyticus.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC Vibrionaceae; Vibrio.

OX NCBI_TaxID=670;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RIMD 2210633 / Serotype O3:k6;

EX MEDLINE=22508454; PubMed=12620739;

RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,

RA Iijima Y., Najima K., Nakano M., Yamashita A., Kubota Y., Kimura S.,

RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;

RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism

RT distinct from that of V. cholerae.";

RL Lancet 361:743-749(2003).

DR EMBL; AP005083; BAC61222.1; -

DR InterPro; IPR002771; MarC.

DR Pfam; PF01914; MarC; 1.

KW Complete proteome.

SQ SEQUENCE 194 AA; 21199 MW; 524EB6BDE9086ADD CRC64;

Query Match 43.8%; Score 7; DB 16; Length 194;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AGQSILS 15

|||||

Db 56 AGQSILS 62

RESULT 3

Q7V4Z6

ID Q7V4Z6 PRELIMINARY; PRT; 194 AA.

AC Q7V4Z6;

DT 01-OCT-2003 (TREMBlrel. 25, Created)

DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Hypothetical protein.

GN PWT1787.

OS Prochlorococcus marinus (strain MIT 9313).

OC Bacteria; Cyanobacteria, Prochlorophytes; Prochlorococcaceae;

OC Prochlorococcus.

OX NCBI_TaxID=74547;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=22825698; PubMed=12917642;

RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,

RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,

RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,

RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,

RA Webb E.A., Zinser E.R., Chisholm S.W.;

RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic

RT niche differentiation.";

RL Nature 424:1042-1047(2003).

DR EMBL; BX572100; CAE21962.1; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 194 AA; 21244 MW; 4B26E976A7ABD593 CRC64;

Query Match 43.8%; Score 7; DB 16; Length 194;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GQSILSG 16

|||||

Db 147 GQSILSG 153

RESULT 4

Q8L356

ID Q8L356 PRELIMINARY; PRT; 258 AA.

AC Q8L356;

DT 01-OCT-2002 (TREMBlrel. 22, Created)

DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)

DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DE Cholera toxin A subunit.

GN CTXA.

OS Vibrio cholerae O27.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC Vibrionaceae; Vibrio.

OX NCBI_TaxID=185331;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=365-96;

EX MEDLINE=21950561; PubMed=11953381;

RA Li M., Shimada T., Morris J.G. Jr., Sulakvelidze A., Sozhamannan S.;

RT "Evidence for the emergence of non-O1 and non-O139 Vibrio cholerae

RT strains with pathogenic potential by exchange of O-antigen

RT biosynthesis regions.";

RL Infect. Immun. 70:2441-2453(2002).

DR EMBL; AF390572; AM22586.1; -

DR GO; GO:0005576; C:extracellular; IEA.

DR GO; GO:0015070; F:toxin activity; IEA.

DR GO; GO:0009405; P:pathogenesis; IEA.

DR InterPro; IPR001144; Enterotoxin A.

DR Pfam; PF01375; Enterotoxin A; 1.

DR PRINTS; PR00771; ENTEROTOXINA.

SQ SEQUENCE 258 AA; 29336 MW; 0F7EBAEE0069ASD0 CRC64;

Query Match 43.8%; Score 7; DB 2; Length 258;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SLRSAHL 8

|||||

Db 83 SLRSAHL 89

RESULT 5

Q8VLI6

ID Q8VLI6 PRELIMINARY; PRT; 258 AA.

AC Q8VLI6;

DT 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE CTXA.

GN CTXA.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC Vibrionaceae; Vibrio.

OX NCBI_TaxID=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=203-93, and 571-88;

RA Li M., Chen Y., Kotetishvili M., Stine O.C., Morris J.G. Jr.,

RA Sulakvelidze A., Sozhamannan S.;

RT "Genetic Analysis of the Virulence Regions, CTX f prophage and Vibrio

RT pathogenicity Island (VPI), in Diverse, Non-epidemic Serogroup Strains

```

RT of Vibrio cholerae.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1322-69;
RA Li M., Chen Y., Kotetishvili M., Stine O.C., Morris J.G. Jr.,
RA Sulakvelidze A., Sozhamannan S.;
RT "Genetic Analysis of the Virulence Regions, CTX f prophage and Vibrio
RT Pathogenicity Island (VPI), in Diverse, Non-epidemic Serogroup Strains
RT of Vibrio cholerae.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF463401; AAL69945.1; -.
DR EMBL; AF452584; AAL60525.1; -.
DR EMBL; AF463400; AAL69944.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR001144; Enterotoxin A.
DR InterPro; IPR000886; ER target S.
DR Pfam; PF01375; Enterotoxin_A; 1.
DR PRINTS; PR00771; ENTEROTOXINA.
DR PROSITE; PS00014; ER_TARGET; 1.
SQ SEQUENCE 258 AA; 29362 MW; 3EA358C7F8BA8BF7 CRC64;

Query Match 43.8%; Score 7; DB 2; Length 258;
Best Local Similarity 100.0%; Pred.No.18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SLRSAHL 8
Db 83 SLRSAHL 89

RESULT 6
Q8LTC8 Q8LTC8 PRELIMINARY; PRT; 258 AA.
AC Q8LTC8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CtxA.
GN CtxA.
OS Vibrio phage CTX.
OC Viruses; ssDNA viruses; Inoviridae; Inovirus.
OX NCBI_TaxID=141904;
RN [1]
RP SEQUENCE FROM N.A.
RA Rui Y., Kan B., Gao S.;
RT "CtxAB gene of strains of Vibrio cholerae isolated from China.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF516341; AM74170.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0015070; F:toxin activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR001144; Enterotoxin A.
DR InterPro; IPR000886; ER target S.
DR Pfam; PF01375; Enterotoxin_A; 1.
DR PRINTS; PR00771; ENTEROTOXINA.
DR PROSITE; PS00014; ER_TARGET; 1.
SQ SEQUENCE 258 AA; 29390 MW; 910FFFE3806B40D6 CRC64;

Query Match 43.8%; Score 7; DB 9; Length 258;
Best Local Similarity 100.0%; Pred.No.18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SLRSAHL 8
Db 83 SLRSAHL 89

RESULT 7
Q7UNSG Q7UNSG PRELIMINARY; PRT; 310 AA.
ID Q7UNSG

```

```

AC Q7UNSG;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN RB7404.
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Firellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Firellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR EMBL; EX294146; CAD75342.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 310 AA; 33824 MW; A8CBFFC1AA4CFAC3 CRC64;

Query Match 43.8%; Score 7; DB 16; Length 310;
Best Local Similarity 100.0%; Pred.No.21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 HLAGQSI 13
Db 63 HLAGQSI 69

RESULT 8
Q8RPJ7 Q8RPJ7 PRELIMINARY; PRT; 335 AA.
AC Q8RPJ7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CprC-like protein.
OS Desulfitobacterium hafnense.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Peptococcaceae;
OC Desulfitobacterium.
OX NCBI_TaxID=49338;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DCB-2;
RA Davis J.K., Tiedje J.M.;
RT "Sequence and transcriptional analysis of reductive dehalogenase genes
RT of Desulfitobacterium.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF403181; AAL87755.1; -.
DR InterPro; IPR007329; FMN bind.
DR Pfam; PF04205; FMN_bind; 1.
SQ SEQUENCE 335 AA; 36945 MW; 5309079265C07E31 CRC64;

Query Match 43.8%; Score 7; DB 2; Length 335;
Best Local Similarity 100.0%; Pred.No.23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SAHLAQ 11
Db 305 SAHLAQ 311

RESULT 9
Q7UHC6 Q7UHC6 PRELIMINARY; PRT; 360 AA.
ID Q7UHC6
AC Q7UHC6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

DE Apatate carbamoyltransferase, catalytic subunit (PyrB) (EC
 DE 2.1.3.2).
 GN PYRB OR RB13301.
 OS Rhodopirellula baltica.
 OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
 OC Planctomycetaceae; Pirellula.
 OX NCBI_TaxID=117;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1;
 RX MEDLINE=22735913; PubMed=12835416;
 RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
 RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
 RA Schlesner H., Amann R., Reinhardt R.;
 RT "Complete genome sequence of the marine planctomycete Pirellula sp.
 RT strain 1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
 DR EMBL; BX294156; CAD78047.1; -;
 KW Transferase; Complete proteome.
 SQ SEQUENCE 360 AA; 39219 MW; 4254BFF1737D5FC4 CRC64;
 Query Match 43.8%; Score 7; DB 16; Length 360;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 LAGOSIL 14
 DB 46 LAGOSIL 52
 RESULT 10
 Q9VIM8 PRELIMINARY; PRT; 561 AA.
 AC Q9VIM8;
 DC 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE CG16798 protein (RE22905p).
 OS CG16798
 GN Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA De Pacios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Stupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RL "The genome sequence of Drosophila melanogaster.";
 RN Science 287:2185-2195 (2000).
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Staponenavong S., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003666; AAF53889.2; -;
 DR EMBL; AY122225; AAM52737.1; -;
 DR FlyBase; FBgn0032856; CG16798.
 DR InterPro; IPR001507; Endoglin/CD105.
 DR Pfam; PF00100; zona_pellucida; 1.
 SQ SEQUENCE 561 AA; 60799 MW; EC717DE4CB96E02D CRC64;
 Query Match 43.8%; Score 7; DB 5; Length 561;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 HLAGOSI 13
 DB 236 HLAGOSI 242


```

DR EMBL; AY071210; AAL48832.1; -.
DR FlyBase; FBgn0010303; hep.
DR GO; GO:0004672; P:protein kinase activity; IDA.
DR GO; GO:0046843; P:eggshell pattern formation (sensu Insecta); IMP.
DR GO; GO:0003081; P:eggshell pattern formation (sensu Insecta); IMP.
DR GO; GO:0000165; P:MAPKK cascade; NAS.
DR GO; GO:0046844; P:micropyle formation; IMP.
DR GO; GO:0007395; P:spreading of leading edge cells; IMP.
SQ SEQUENCE 580 AA; 61880 MW; 3596898A7A7A5F9F CRC64;

Query Match 43.8%; Score 7; DB 5; Length 580;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LAGQSIL 14
DB 52 LAGQSIL 58

RESULT 13
Q917S3 PRELIMINARY; PRT; 580 AA.
ID Q917S3;
AC Q917S3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG4353 protein.
DE HEP OR CG2190 OR CG4353.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adamantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

```

RT "The genome sequence of *Drosophila melanogaster*.";

RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Celisner S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,

RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hoetin D., Howland T.J.,

RA Ibegwam C., Jalali M., Kursey D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson K.A., Nunoo J.,

RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,

RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;

RT "Sequencing of *Drosophila melanogaster* genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Mara S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kamniker S.E., Prochnik S.E., Smith C.D.,

RA Tupy J.L., Bergman C., Bertram B., Carlson J.W., Celisner S.E.,

RA Clapp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,

RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,

RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;

RT "Annotation of *Drosophila melanogaster* genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Adams M.D., Celisner S.E., Gibbs R.A., Rubin G.M., Venter C.J.;

RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RL [5]

RP SEQUENCE FROM N.A.

RA FlyBase;

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE003491; AAG22351.2; --

DR FlyBase; PBgn0010303; Rep.

DR GO; GO:0004672; P:protein kinase activity; IDA.

DR GO; GO:0046843; P:dorsal appendage formation; IMP.

DR GO; GO:0030381; P:eggshell pattern formation (sensu Insecta); IMP.

DR GO; GO:0000165; P:MAPKKK cascade; NAS.

DR GO; GO:0000165; P:microtubule formation; IMP.

DR GO; GO:0046844; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkman L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,
RA Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoef A., Tuemmler B.,
RA Fraser C.M.;

RT "Complete genome sequence and comparative analysis of the

RT metabolically versatile *Pseudomonas putida* KT2440.";

RL Environ. Microbiol. 4:1799-808(2002).

DR EMBL; AE016792; AAN70342.1; --

DR TIGR; PP4772; --

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.

DR GO; GO:0003876; F:nucleic acid binding; IEA.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR001650; Helicase_C.

DR InterPro; IPR007502; Helicase_dom.

DR Pfam; PF04408; HA2; 1.

DR Pfam; PF00271; helicase C; 1.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

DR Helicase; Complete proteome.

RT "The genome sequence of *Drosophila melanogaster*.";

RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Celisner S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,

RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hoetin D., Howland T.J.,

RA Ibegwam C., Jalali M., Kursey D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson K.A., Nunoo J.,

RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,

RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;

RT "Sequencing of *Drosophila melanogaster* genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Mara S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kamniker S.E., Prochnik S.E., Smith C.D.,

RA Tupy J.L., Bergman C., Bertram B., Carlson J.W., Celisner S.E.,

RA Clapp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,

RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,

RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;

RT "Annotation of *Drosophila melanogaster* genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Adams M.D., Celisner S.E., Gibbs R.A., Rubin G.M., Venter C.J.;

RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RL [5]

RP SEQUENCE FROM N.A.

RA FlyBase;

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE003491; AAG22351.2; --

DR FlyBase; PBgn0010303; Rep.

DR GO; GO:0004672; P:protein kinase activity; IDA.

DR GO; GO:0046843; P:dorsal appendage formation; IMP.

DR GO; GO:0030381; P:eggshell pattern formation (sensu Insecta); IMP.

DR GO; GO:0000165; P:MAPKKK cascade; NAS.

DR GO; GO:0000165; P:microtubule formation; IMP.

DR GO; GO:0046844; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading of leading edge cells; IMP.

DR GO; GO:0007395; P:spreading

Fri Apr 23 15:38:29 2004

us-09-528-682-1_copy_64_79.oligo.rspt

Page 7

Search completed: April 23, 2004, 14:54:07
Job time : 35.3333 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:45:50 ; Search time 48 Seconds
(without alignments)

94.183 Million cell updates/sec

Title: US-09-528-682-1_COPY_64_79

Perfect score: 16

Sequence: 1 LSLRSHLAGQSILSG 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq_29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1980s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	100.0	237	5	Abb07780 E coli mu
2	16	100.0	240	2	Aaw65074 E. coli L
3	16	100.0	240	5	Abb07778 E coli he
4	16	100.0	258	1	Aap50190 Sequence
5	16	100.0	258	1	Aap50191 Sequence
6	16	100.0	259	3	Aay96647 Synthetic
7	16	100.0	259	3	Aay96648 Plant-opt
8	16	100.0	259	3	Aay96646 Plant-opt
9	16	100.0	259	3	Aay96650 Plant-opt
10	16	100.0	370	5	Abb07785 E coli mu
11	16	100.0	373	5	Abb07784 E coli he
12	16	100.0	380	4	Aau00507 E. coli h
13	16	100.0	382	4	Aau00506 E. coli h
14	11	68.8	236	2	Aar38728 E.coli he
15	11	68.8	236	2	Aar44024 "Glu-114"
16	11	68.8	236	2	Aar44023 "Ser-106"
17	11	68.8	236	2	Aar44022 "Ser-104"
18	11	68.8	236	2	Aar44021 "Asp-53"
19	11	68.8	236	2	Aar38730 "Lys-104"
20	11	68.8	236	2	Aar44020 "Tyr-53"
21	11	68.8	236	2	Aar38732 "Lys-114"
22	11	68.8	236	2	Aar44025 "Glu-107"
23	11	68.8	236	2	Aar44019 "Lys-97"
24	11	68.8	236	2	Aar44017 "Tyr-97"
25	11	68.8	236	2	Aar44018 "Tyr-97"

ALIGNMENTS

RESULT 1

ABB07780

ID ABB07780 standard; protein; 237 AA.

AC ABB07780;

DT 17-JUN-2002 (first entry)

DE E coli mutant heat-labile toxin (mLT) A subunit fragment.

XX LT; heat-labile toxin; cholera toxin; CT; recombinant; adjuvant; vaccine;
mutant; mLT.

OS Escherichia coli.

XX JP2002051779-A.

XX 19-FEB-2002.

XX 07-AUG-2000; 2000JP-00238740.

XX 07-AUG-2000; 2000JP-00238740.

XX (DOKU-) DOKURITSU GYOSEI HOJIN NOGYO SEIBUTSU SH.

XX (HGET) HIGETA SHOYU KK.

XX (FUJI-) FUJITA GAKUEN.

XX WPI; 2002-299402/34.

XX Preparation of a protein having 1A5B structure.

XX Claim 6; Fig 3; 27pp; Japanese.

XX The invention relates to a gene encoding a protein having a subunit structure of 1A5B in which the DNA sequence encoding each signal is deleted from the A subunit gene and the B subunit gene and they are combined tandemly in the order of (B subunit gene)-(SD sequence gene)-(A subunit gene). A method is provided for the preparation of a protein having a subunit structure of 1A5B in which the above DNA is connected to a vector expressible in Brevibacillus chosinensis and Brevibacillus chosinensis is transformed by said vector and said transformant is cultured. The protein can be used in the preparation of an adjuvant for vaccine. The present sequence represents the E. coli mutant heat-labile toxin (mLT) A subunit fragment, used for constructing a protein with the subunit structure of 1A5B

XX Sequence 237 AA;

SQ

Aar38731 "Glu-53"
Aar44016 "Lys-63"
Aaw67772 E. coli h
Aau14105 Peptide s
Abo10270 E. coli h
Aay96649 Plant-opt
Aay96651 Plant-opt
Aar72526 ADP-ribos
Aay41797 ADP-ribos
Aaw95207 Peptide f
Aay68346 Heat labi
Aab66220 E coli he
Adc96083 E. faeciu
Aar44033 Glu-114 c
Aar44038 Lys-97 ch
Aar44030 An-107 c
Aar44029 Ser-106 c
Aar44031 Ser-110 c
Aar38729 Cholera t

26 11 68.8 236 2 AAR38731
27 11 68.8 236 2 AAR44016
28 11 68.8 237 2 AAW67772
29 11 68.8 254 4 AAU14105
30 11 68.8 254 6 ABO10270
31 8 50.0 240 2 AAW65075
32 8 50.0 259 3 AAY96649
33 8 50.0 259 3 AAY96651
34 7 43.8 12 2 AAR72526
35 7 43.8 12 2 AAY41797
36 7 43.8 12 2 AAW95207
37 7 43.8 12 2 AAY68346
38 7 43.8 12 3 AAB66220
39 7 43.8 215 7 ADC96083
40 7 43.8 240 2 AAR44033
41 7 43.8 240 2 AAR44028
42 7 43.8 240 2 AAR44030
43 7 43.8 240 2 AAR44029
44 7 43.8 240 2 AAR44031
45 7 43.8 240 2 AAR38729

Query Match 100.0%; Score 16; DB 5; Length 237;
 Best Local Similarity 100.0%; Pred. No. 1.3e-08;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSALHLAGQSILSG 16
 |||||
 Db 64 LSLRSALHLAGQSILSG 79

RESULT 2
 AAW65074
 ID AAW65074 standard; protein; 240 AA.
 XX
 AC AAW65074;
 XX
 DT 11-SEP-1998 (first entry)
 XX
 DE E. coli LT-A protein fragment.
 XX
 KW Heat labile toxin subunit A; LT-A; mutant; immunogen; detoxification;
 carrier; adjuvant; prevention; treatment; disease; diarrhoea; vaccine;
 infection; enterotoxic.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT Protein 1..241
 FT /note= "partial sequence"
 XX
 PN WO9818928-A1.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-IB001440.
 XX
 PR 31-OCT-1996; 96GB-00022660.
 XX
 PA (CHIR-) CHIRON SPA.
 XX
 PI Pizza M, Giuliani MM, Rappuoli R;
 XX
 DR WPI; 1998-272223/24.
 XX
 PT Mutated Escherichia coli heat labile toxin subunit A - is immunogenic and
 detoxified relative to wild-type, useful e.g. in vaccines against E. coli
 enterotoxigenic strains and as an adjuvant.
 XX
 PS Disclosure; Page: 67pp; English.

This sequence represents a fragment of an Escherichia coli heat labile
 toxin subunit A (LT-A). This protein is used in a method resulting in a
 mutant LT-A protein which has the wild type Ala residue at position 72
 changed to an Arg residue resulting in a toxin which retains its
 immunogenicity but is detoxified. Detoxification is defined in the
 specification as a reduction in toxicity relative to wild-type toxin,
 such that any residual toxicity is sufficiently low to allow use as an
 effective immunogenic composition in humans without significant side
 effects. The protein can be combined with an acceptable carrier in
 immunogenic compositions, optionally comprising an adjuvant and/or a
 second immunogenic antigen. Such compositions can be administered to
 prevent/treat disease in a subject e.g. traveller's diarrhoea in humans.
 The protein or compositions are especially administered as vaccines
 useful to prevent or treat infections by enterotoxigenic strain of E.
 coli in mammals (especially humans)

XX Sequence 240 AA;
 Query Match 100.0%; Score 16; DB 2; Length 240;
 Best Local Similarity 100.0%; Pred. No. 1.3e-08;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSALHLAGQSILSG 16
 |||||
 Db 64 LSLRSALHLAGQSILSG 79

Db 64 LSLRSALHLAGQSILSG 79

RESULT 3
 ABB07778
 ID ABB07778 standard; protein; 240 AA.
 XX
 AC ABB07778;
 XX
 DT 17-JUN-2002 (first entry)
 XX
 DE E coli heat-labile toxin (LT) A subunit fragment.
 XX
 KW LT; heat-labile toxin; cholera toxin; CT; recombinant; adjuvant; vaccine.
 XX
 OS Escherichia coli.
 XX
 PN JP2002051779-A.
 XX
 PD 19-FEB-2002.
 XX
 PF 07-AUG-2000; 2000JP-00238740.
 XX
 PR 07-AUG-2000; 2000JP-00238740.
 XX
 PA (DOKU-) DOKURITSU GYOSEI HOJIN NOGYO SEIBUTSU SH.
 (HGET) HIGETA SHOYU KK.
 (FUJI-) FUJITA GAKUEN.
 XX
 DR WPI; 2002-299402/34.
 XX
 PT Preparation of a protein having 1A5B structure.
 XX
 PS Claim 5; Fig 1; 27pp; Japanese.
 XX
 CC The invention relates to a gene encoding a protein having a subunit
 structure of 1A5B in which the DNA sequence encoding each signal is
 deleted from the A subunit gene and the B subunit gene and they are
 combined tandemly in the order of (B subunit gene)-(SD sequence gene)-(A
 subunit gene). A method is provided for the preparation of a protein
 having a subunit structure of 1A5B in which the above DNA is connected to
 a vector expressible in Brevibacillus chosinensis and Brevibacillus
 chosinensis is transformed by said vector and said transformant is
 cultured. The protein can be used in the preparation of an adjuvant for
 vaccine. The present sequence represents the E. coli heat-labile toxin
 (LT) A subunit fragment, used for constructing a protein with the subunit
 structure of 1A5B

XX Sequence 240 AA;
 Query Match 100.0%; Score 16; DB 5; Length 240;
 Best Local Similarity 100.0%; Pred. No. 1.3e-08;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSALHLAGQSILSG 16
 |||||
 Db 64 LSLRSALHLAGQSILSG 79

RESULT 4
 AAP50190
 ID AAP50190 standard; protein; 258 AA.
 XX
 AC AAP50190;
 XX
 DT 27-AUG-2003 (revised)
 DT 30-OCT-1991 (first entry)
 XX
 DE Sequence encoded by the pig scours heat labile toxin (LT) LTA gene.
 XX
 KW Pig scours vaccine; toxin; diarrhoea.
 XX
 OS Escherichia coli.

```

XX EP145486-A.
PN 19-JUN-1985.
XX
XX 12-DEC-1984; 84EP-00308620.
XX
XX 12-DEC-1983; 83GB-00033131.
XX
XX (GLAX ) GLAXO GROUP LTD.
XX
XX Hayes MV, Harford S, Ross GW;
XX
XX WPI; 1985-148358/25.
XX
XX N-PSDB; AAN50205.
XX
XX New toxoid as inactivated form of toxin for use in vaccines - is obt'd.
XX from organism transformed by gene.
XX
XX Disclosure; Fig 1; 61pp; English.
XX
XX AAN50205 is the gene sequence of the natural LTA gene. The LTA gene of
XX the site directed mutant SDM1 (see AAN50206) is inactive. The inventors
XX claim a vaccine prepn. active against pig scours which contains an
XX inactivated LTA component, together with additional K88 antigens opt.
XX with whole cells comprising the antigens or contg. the inactivated LTA.
XX (Updated on 27-AUG-2003 to correct OS field.)
XX
XX Sequence 258 AA;
XX
Query Match 100.0%; Score 16; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLSRAHLAQSIILSG 16
Db 82 LSLSRAHLAQSIILSG 97

RESULT 5
AAP50191
ID AAP50191 standard; protein; 258 AA.
XX
AC AAP50191;
XX
XX 27-AUG-2003 (revised)
DT 30-OCT-1991 (first entry)
XX
XX Sequence encoded by the pig scours heat labile toxin (LT) LTA gene of the
XX site directed mutant SDM1.
XX
XX Pig scours vaccine; toxin; diarrhoea.
XX
XX Escherichia coli.
XX
XX Key Location/Qualifiers
XX Misc-difference 79 /note= "Ser in native SQ"
XX
XX EP145486-A.
XX
XX 19-JUN-1985.
XX
XX 12-DEC-1984; 84EP-00308620.
XX
XX 12-DEC-1983; 83GB-00033131.
XX
XX (GLAX ) GLAXO GROUP LTD.
XX
XX Hayes MV, Harford S, Ross GW;
XX
XX WPI; 1985-148358/25.
XX
XX N-PSDB; AAN50206.

```

```

XX
XX New toxoid as inactivated form of toxin for use in vaccines - is obt'd.
XX from organism transformed by gene.
XX
XX Example; Fig 2; 61pp; English.
XX
XX AAN50205 is the gene sequence of the natural LTA gene. The LTA gene of
XX the site directed mutant SDM1 (see AAN50206) is inactive. The inventors
XX claim a vaccine prepn. active against pig scours which contains an
XX inactivated LTA component, together with additional K88 antigens opt.
XX with whole cells comprising the antigens or contg. the inactivated LTA.
XX (Updated on 27-AUG-2003 to correct OS field.)
XX
XX Sequence 258 AA;
XX
Query Match 100.0%; Score 16; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLSRAHLAQSIILSG 16
Db 82 LSLSRAHLAQSIILSG 97

RESULT 6
AAV96647
ID AAV96647 standard; protein; 259 AA.
XX
XX AAV96647;
XX
XX 26-SEP-2000 (first entry)
DT
XX
XX Synthetic E. coli LT-A K63 mutant.
DE
XX
XX Heat-labile toxin; LT-A; LT-B; mutein; transgenic plant; vaccine; oral;
XX adjuvant; anti-bacterial; S63K.
XX
XX Escherichia coli.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Peptide 1..19 /label= signal_peptide
XX /note= "20"
XX
XX Protein 20..259 /label= mature_protein
XX
XX Misc-difference 82 /label= S63K
XX /note= "Wild-type serine is replaced by lysine"
XX
XX WO200037609-A2.
XX
XX 29-JUN-2000.
XX
XX 22-DEC-1999; 99WO-US030747.
XX
XX 22-DEC-1998; 98US-0113507P.
XX
XX (BOYC-) BOYCE THOMPSON INST PLANT RES.
XX (MASO/) MASON H S.
XX (ARNT/) ARNTZEN C J.
XX
XX Mason HS, Arntzen CJ;
XX
XX WPI; 2000-442653/38.
XX
XX N-PSDB; AAS1147.
XX
XX New polynucleotides encoding LT-A or CT-A polypeptides for the
XX transformation of plant cells, useful in immunogenic compositions to
XX elicit immune responses in animals.
XX
XX Example 2; Page; 103pp; English.
XX
XX

```

CC This is mutant S63K Escherichia coli heat-labile toxin (LT) A subunit (LT
CC -A). The wild-type serine was replaced with lysine at residue 63 of the
CC mature protein, which was caused by a codon change of TCC to AAG in the
CC coding sequence. The sequence contains plant-preferred codons and
CC eliminates sequence motifs associated with spurious mRNA processing. A
CC single codon insertion (GTG encoding valine) was made to accommodate the
CC creation of a NcoI restriction site around the initiator methionine
CC codon. Novel polynucleotides encode a mutant LT-A polypeptide or a mutant
CC Vibrio cholerae cholera toxin (CT) A subunit (CT-A) polypeptide, which
CC have reduced enzyme activity as compared to the wild-type LT-A or CT-A
CC polypeptide and where at least one of the codons is altered to a plant
CC preferred codon. The polynucleotide further comprises a nucleic acid
CC sequence encoding LT B subunit (LT-B) or a CT B subunit (CT-B). The
CC polynucleotides are useful for the transformation of plant cells for the
CC production of transgenic plants to produce edible vaccines, especially
CC oral vaccines in transgenic plants for the prophylactic or therapeutic
CC treatment against E. coli or V. cholerae. The mutant polypeptides are
CC also useful as adjuvants. NB: This sequence does not appear in the
CC specification, it was made from the wild type sequence shown in AAY96647,
CC which appears in Figure 1

XX Sequence 259 AA;

Query Match 100.0%; Score 16; DB 3; Length 259;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LSLRSAHLAQSTLSG 16
DB 83 LSLRSAHLAQSTLSG 98

RESULT 7
AAY96648
ID AAY96648 standard; protein; 259 AA.

XX AAY96648;

XX 26-SEP-2000 (first entry)

XX Plant-optimized E. coli LT-A G192 mutant.

XX Heat-labile toxin; LT-A; LT-B; mutant; transgenic plant; vaccine; oral;
XX adjuvant; anti-bacterial; R192G.

XX Escherichia coli.

XX Synthetic.

XX Key Location/Qualifiers
XX Peptide 1..19
XX /label= signal_peptide
XX Protein 20..259
XX /label= mature_protein
XX Misc-difference 211
XX /label= R192G
XX /note= "Wild-type arginine is replaced by glycine"

XX WO200037609-A2.

XX 29-JUN-2000.

XX 22-DEC-1999; 99WO-US030747.

XX 22-DEC-1998; 98US-0113507P.

XX (BOYC-) BOYCE THOMPSON INST PLANT RES.
XX PA (MASO/) MASON H S.
XX PA (ARNT/) ARNTZEN C J.

XX Mason HS, Arntzen CJ;

XX WPI; 2000-442653/38.

XX N-PSDB; AAA51544.

XX New polynucleotides encoding LT-A or CT-A polypeptides for the
XX transformation of plant cells; useful in immunogenic compositions to
XX elicit immune responses in animals.

XX Example 3; Page; 103pp; English.

XX This is mutant R192G Escherichia coli heat-labile toxin (LT) A subunit
XX (LT-A). The wild-type arginine was replaced with glycine at residue 192
XX of the mature protein, which was caused by a codon change of TCC to AAG
XX in the coding sequence. The sequence contains plant-preferred codons and
XX eliminates sequence motifs associated with spurious mRNA processing. A
XX single codon insertion (GTG encoding valine) was made to accommodate the
XX creation of a NcoI restriction site around the initiator methionine
XX codon. Novel polynucleotides encode a mutant LT-A polypeptide or a mutant
XX Vibrio cholerae cholera toxin (CT) A subunit (CT-A) polypeptide, which
XX have reduced enzyme activity as compared to the wild-type LT-A or CT-A
XX polypeptide and where at least one of the codons is altered to a plant
XX preferred codon. The polynucleotide further comprises a nucleic acid
XX sequence encoding LT B subunit (LT-B) or a CT B subunit (CT-B). The
XX polynucleotides are useful for the transformation of plant cells for the
XX production of transgenic plants to produce edible vaccines, especially
XX oral vaccines in transgenic plants for the prophylactic or therapeutic
XX treatment against E. coli or V. cholerae. The mutant polypeptides are
XX also useful as adjuvants. NB: This sequence does not appear in the
XX specification, it was made from the wild type sequence shown in AAY96647,
XX which appears in Figure 1

XX Sequence 259 AA;

Query Match 100.0%; Score 16; DB 3; Length 259;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LSLRSAHLAQSTLSG 16
DB 83 LSLRSAHLAQSTLSG 98

RESULT 8

AAY96646

ID AAY96646 standard; protein; 259 AA.

XX AAY96646;

XX 26-SEP-2000 (first entry)

XX Plant-optimized E. coli heat labile toxin A subunit.

XX Heat-labile toxin; LT-A; LT-B; mutant; transgenic plant; vaccine; oral;
XX adjuvant; anti-bacterial.

XX Escherichia coli.

XX Synthetic.

XX Key Location/Qualifiers
XX Peptide 1..18
XX /label= signal_peptide
XX Protein 19..259
XX /label= mature_protein

XX WO200037609-A2.

XX 29-JUN-2000.

XX 22-DEC-1999; 99WO-US030747.

XX 22-DEC-1998; 98US-0113507P.

XX (BOYC-) BOYCE THOMPSON INST PLANT RES.
XX PA (MASO/) MASON H S.
XX PA (ARNT/) ARNTZEN C J.

22-DEC-1999; 99WO-US030747.
22-DEC-1998; 98US-0113507P.
(BOYC-) BOYCE THOMPSON INST PLANT RES.
(MASO/) MASON H S.
(ARNT/) ARNTZEN C J.
Mason HS, Arntzen CJ;
WPI; 2000-442653/38.
New polynucleotides encoding LT-A or CT-A polypeptides for the transformation of plant cells, useful in immunogenic compositions to elicit immune responses in animals.
Example 5; Page; 103pp; English.
This is Escherichia coli heat-labile toxin (LT) A subunit (LT-A) double mutant S63xR192G. The wild-type serine was replaced with lysine at residue 63 and wild-type arginine was replaced with glycine at residue 192 of the mature protein. The coding sequence contains plant-preferred codons and eliminates sequence motifs associated with spurious mRNA processing. A single codon insertion (GTG encoding valine) was made to accommodate the creation of a NcoI restriction site around the initiator methionine codon. Novel polynucleotides encode a mutant LT-A polypeptide or a mutant Vibrio cholerae cholera toxin (CT) A subunit (CT-A) polypeptide, which have reduced enzyme activity as compared to the wild-type LT-A or CT-A polypeptide and where at least one of the codons is altered to a plant preferred codon. The polynucleotide further comprises a nucleic acid sequence encoding LT B subunit (LT-B) or a CT B subunit (CT-B). The polynucleotides are useful for the transformation of plant cells for the production of transgenic plants to produce edible vaccines, especially oral vaccines in transgenic plants for the prophylactic or therapeutic treatment against E. coli or V. cholerae. The mutant polypeptides are also useful as adjuvants. Note: This sequence does not appear in the specification. It was constructed from the wild type LT-A shown in AAY96646 which is given in Figure 1 of the specification
Sequence 259 AA;
Query Match 100.0%; Score 16; DB 3; Length 259;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLSRSAHLAGQSILSG 16
|||
DB 83 LLSRSAHLAGQSILSG 98
|||

RESULT 10
ABB07785
ID ABB07785 standard; protein; 370 AA.
XX AC ABB07785;
XX AC
XX AC
DT 17-JUN-2002 (first entry)
DE DE
XX KW E coli mutant heat-labile toxin (mLT) 5B-SD-1A protein.
KW KW LT; heat-labile toxin; cholera toxin; CT; recombinant; adjuvant; vaccine;
KW KW mutant; mLT.
OS Escherichia coli.
XX Key Location/Qualifiers
XX Peptide 1..103
FT FT /note= "LT 5B subunit"
FT FT 104..370
FT FT /note= "LT 1A subunit"
XX XX
XX JF2002051779-A.
XX FN
XX XX

PD 19-FEB-2002.
 XX 07-AUG-2000; 2000JP-00238740.
 XX 07-AUG-2000; 2000JP-00238740.
 XX (DOKU-) DOKURITSU GYOSEI HOJIN NOGYO SEIBUTSU SH.
 PA (HGET) HIGETA SHOYU KK.
 PA (FUJI-) FUJITA GAKUEN.
 XX WPI; 2002-299402/34.
 DR N-PSDB; ABL40640.
 XX Preparation of a protein having 1A5B structure.
 PT Disclosure; Fig 9-10; 27pp; Japanese.
 XX The invention relates to a gene encoding a protein having a subunit structure of 1A5B in which the DNA sequence encoding each signal is deleted from the A subunit gene and the B subunit gene and they are combined tandemly in the order of (B subunit gene)-(SD sequence gene)-(A subunit gene). A method is provided for the preparation of a protein having a subunit structure of 1A5B in which the above DNA is connected to a vector expressible in *Brevibacillus chosinensis* and *Brevibacillus chosinensis* is transformed by said vector and said transformant is cultured. The protein can be used in the preparation of an adjuvant for vaccine. The present sequence represents the *E. coli* mutant heat-labile toxin (mLT) 5B-SD-1A protein

Query Match 100.0%; Score 16; DB 5; Length 370;
 Best Local Similarity 100.0%; Pred. No. 1.9e-08; Indels 0; Gaps 0;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LSLSRAHLAQSIILSG 16
 |||||
 Db 197 LSLSRAHLAQSIILSG 212
 |||||

RESULT 11
 ABB07784
 ID ABB07784 standard; protein; 373 AA.
 XX ABB07784;
 AC ABB07784;
 XX 17-JUN-2002 (first entry)
 DT LT; heat-labile toxin (LT) 5B-SD-1A protein.
 DE LT; heat-labile toxin; cholera toxin; CT; recombinant; adjuvant; vaccine.
 XX *Escherichia coli*.
 OS
 XX Key Location/Qualifiers
 FH Peptide 1..103
 FT Peptide /note= "LT 5B subunit"
 FT Peptide 104..373
 FT Peptide /note= "LT 1A subunit"
 XX JP2002051779-A.
 PN 19-FEB-2002.
 PD 07-AUG-2000; 2000JP-00238740.
 PF 07-AUG-2000; 2000JP-00238740.
 XX (DOKU-) DOKURITSU GYOSEI HOJIN NOGYO SEIBUTSU SH.
 PA (HGET) HIGETA SHOYU KK.
 PA (FUJI-) FUJITA GAKUEN.
 XX WPI; 2002-299402/34.
 DR

DR N-PSDB; ABL40639.
 XX Preparation of a protein having 1A5B structure.
 PT Disclosure; Fig 7-8; 27pp; Japanese.
 XX The invention relates to a gene encoding a protein having a subunit structure of 1A5B in which the DNA sequence encoding each signal is deleted from the A subunit gene and the B subunit gene and they are combined tandemly in the order of (B subunit gene)-(SD sequence gene)-(A subunit gene). A method is provided for the preparation of a protein having a subunit structure of 1A5B in which the above DNA is connected to a vector expressible in *Brevibacillus chosinensis* and *Brevibacillus chosinensis* is transformed by said vector and said transformant is cultured. The protein can be used in the preparation of an adjuvant for vaccine. The present sequence represents the *E. coli* heat-labile toxin (LT) 5B-SD-1A protein

Query Match 100.0%; Score 16; DB 5; Length 373;
 Best Local Similarity 100.0%; Pred. No. 1.9e-08; Indels 0; Gaps 0;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LSLSRAHLAQSIILSG 16
 |||||
 Db 197 LSLSRAHLAQSIILSG 212
 |||||

RESULT 12
 AAU00507
 ID AAU00507 standard; protein; 380 AA.
 XX AAU00507;
 AC AAU00507;
 XX 29-AUG-2001 (first entry)
 DT *E. coli* heat-labile enterotoxin (LT) mutant LTdel110/112.
 DE Heat-labile enterotoxin; LT; LTSE63Y; LTdel110/112; mutant; detoxified and immunologically active protein; ADP-ribosylation; Gs; endotoxin; diarrhoea; mutein.
 KW *Escherichia coli*; strain K88ac.
 OS Synthetic.
 XX Key Location/Qualifiers
 FH Misc_feature 7
 FT /note= "Important residue for enzymatic activity"
 FT Misc_feature 44
 FT /note= "Important residue for enzymatic activity"
 FT Region 58..72
 FT /note= "Forms the NAD-binding site"
 FT Misc_feature 61
 FT /note= "Important residue for enzymatic activity"
 FT Misc-difference 257
 FT /note= "Encoded by TG"
 XX WO200119998-A1.
 PN 22-MAR-2001.
 PD 15-SEP-1999; 99WO-KR000555.
 PF 15-SEP-1999; 99WO-KR000555.
 XX 15-SEP-1999; 99WO-KR000555.
 PR (MOGA-) MOGAM BIOTECHNOLOGY RES INST.
 XX Park EJ, Kim JS, Chang J, Yum J, Chung S;
 PI WPI; 2001-281524/29.
 DR N-PSDB; AAS01506.
 XX

PT New detoxified mutants of Escherichia coli heat-labile enterotoxin useful
 PT as vaccine for preventing and treating diarrhea, and as adjuvant for
 PT antibody production.
 XX
 XX
 PS Claim 6; Page 42-44; 48pp; English.
 XX
 XX The present sequence represents Escherichia coli heat-labile enterotoxin
 CC (LT) mutant L7del110/112. L7S63Y (AAU00506) and L7del110/112 are two
 CC novel detoxified and immunologically active proteins (LT mutants) derived
 CC by site-directed mutagenesis of the A1 subunit of wild type LT. The
 CC substitution of Ser to Tyr at position 63 in L7S63Y blocks NAD-binding.
 CC Deletion of Glu residues at positions 110 and 112 in L7del110/112
 CC eliminate the enzymatic activity of LT. The A1 subunit of wild type LT
 CC catalyzes ADP-ribosylation of Gs, a GTP-binding protein that regulates
 CC CAMP levels. The resulting increase in cAMP is the cause of diarrhea in
 CC humans and animals e.g. pigs. The mucosal immunogenicities of mutant heat
 CC -labile endotoxins L7S63Y and L7del110/112 were tested. Groups of mice
 CC were immunised with L7S63Y or L7del110/112. The control groups received
 CC phosphate buffered saline (PBS) alone. The serum and faecal antibody
 CC titres to LT were determined. The results showed that mice immunised with
 CC L7S63Y or L7del110/112 contained high and comparable level of anti-LT
 CC antibodies in sera and faecal extracts compared with those immunised with
 CC wild-type LT. The LT mutants are useful as a vaccine for preventing and
 CC treating diarrhea and as an adjuvant for antibody production
 XX
 SQ Sequence 380 AA;
 Query Match 100.0%; Score 16; DB 4; Length 380;
 Best Local Similarity 100.0%; Pred. No. 2e-08;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LLSRSHLAGQSILSG 16
 |||||
 DB 82 LLSRSHLAGQSILSG 97
 |||||
 RESULT 13
 AAU00506
 ID AAU00506 standard; protein; 382 AA.
 AC AAU00506;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE E. coli heat-labile enterotoxin (LT) mutant L7S63Y.
 XX
 KW Heat-labile enterotoxin; LT; L7S63Y; L7del110/112; mutant;
 KW detoxified and immunologically active protein; ADP-ribosylation; Gs;
 KW endotoxin; diarrhea; mitein.
 XX
 OS Escherichia coli; strain K88ac.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc_feature 7
 FT /note= "Important residue for enzymatic activity"
 FT Misc_feature 44
 FT /note= "Important residue for enzymatic activity"
 FT Region 58..72
 FT /note= "Forms the NAD-binding site"
 FT Misc_feature 61
 FT /note= "Important residue for enzymatic activity"
 FT Misc-difference 63
 FT /note= "Substitution of wild type Ser to Tyr"
 FT Misc_feature 110
 FT /note= "Important residue for enzymatic activity"
 FT Misc_feature 112
 FT /note= "Important residue for enzymatic activity"
 FT Misc-difference 259
 FT /note= "Encoded by TG"
 FT
 XX WO200119998-A1.
 PN
 XX

PD 22-MAR-2001.
 XX
 PF 15-SEP-1999; 99WO-KR000555.
 XX
 PR 15-SEP-1999; 99WO-KR000555.
 XX
 PA (MOGA-) MOGAM BIOTECHNOLOGY RES INST.
 XX
 PI Park EJ, Kim JS, Chang J, Yum J, Chung S;
 DR WPI; 2001-281524/29.
 DR N-PSDB; AAS01505.
 XX
 PT New detoxified mutants of Escherichia coli heat-labile enterotoxin useful
 PT as vaccine for preventing and treating diarrhea, and as adjuvant for
 PT antibody production.
 XX
 PS Claim 2; Page 39-41; 48pp; English.
 XX
 CC The present sequence represents Escherichia coli heat-labile enterotoxin
 CC (LT) mutant L7S63Y. L7S63Y and L7del110/112 (AAU00507) are two novel
 CC detoxified and immunologically active proteins (LT mutants) derived by
 CC site-directed mutagenesis of the A1 subunit of wild type LT. The
 CC substitution of Ser to Tyr at position 63 in L7S63Y blocks NAD-binding.
 CC Deletion of Glu residues at positions 110 and 112 in L7del110/112
 CC eliminate the enzymatic activity of LT. The A1 subunit of wild type LT
 CC catalyzes ADP-ribosylation of Gs, a GTP-binding protein that regulates
 CC CAMP levels. The resulting increase in cAMP is the cause of diarrhea in
 CC humans and animals e.g. pigs. The mucosal immunogenicities of mutant heat
 CC -labile endotoxins L7S63Y and L7del110/112 were tested. Groups of mice
 CC were immunised with L7S63Y or L7del110/112. The control groups received
 CC phosphate buffered saline (PBS) alone. The serum and faecal antibody
 CC titres to LT were determined. The results showed that mice immunised with
 CC L7S63Y or L7del110/112 contained high and comparable level of anti-LT
 CC antibodies in sera and faecal extracts compared with those immunised with
 CC wild-type LT. The LT mutants are useful as a vaccine for preventing and
 CC treating diarrhea and as an adjuvant for antibody production
 XX
 SQ Sequence 382 AA;
 Query Match 100.0%; Score 16; DB 4; Length 382;
 Best Local Similarity 100.0%; Pred. No. 2e-08;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LLSRSHLAGQSILSG 16
 |||||
 DB 82 LLSRSHLAGQSILSG 97
 |||||
 RESULT 14
 AAR38728
 ID AAR38728 standard; protein; 236 AA.
 AC AAR38728;
 XX
 DT 25-MAR-2003 (revised)
 DT 08-DEC-1993 (first entry)
 XX
 DE E. coli heat labile toxin subunit A.
 XX
 KW enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
 KW protomer A; site-directed mutagenesis; reduced toxicity;
 KW ADP-ribosyltransferase activity.
 XX
 OS Escherichia coli.
 XX
 PN WO9313202-A1.
 XX
 PD 08-JUL-1993.
 XX
 PF 30-DEC-1992; 92WO-EF003016.
 XX
 PF 31-DEC-1991; 91IT-MT003513.
 XX

```

XX PA (BIOC-) BIOGINE SCLAVO SPA.
XX PI Domenighini M, Rappuoli R, Pizza M, Hol W;
XX XX
XX WPI; 1993-227320/28.
XX DR N-PSDB; AAQ42768.
XX XX
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin -
PT useful as vaccines against infection by Vibrio cholerae and enterotoxin
PT producing Escherichia coli.
XX XX
XX Disclosure; Fig 2; 60pp; English.
XX CC
XX CC This is the sequence of the A subunit of the heat labile toxin (LT-A) of
CC a strain of E.coli known to affect humans. The sequence was published by
CC Yamamoto et al, J.Biol. Chem., 259, 5037-5044. Mutations at selected
CC positions within this sequence have been found to reduce toxicity (see
CC AAR38730-R38732 and AAR44016-R44025). The invention relates to such
CC immunogenic, detoxified proteins and their use in vaccines to protect
CC against enterotoxigenic E.coli. (Updated on 25-MAR-2003 to correct PN
CC field.) (Updated on 25-MAR-2003 to correct PR field.)
XX SQ Sequence 236 AA;

Query Match 68.8%; Score 11; DB 2; Length 236;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSAHLAQ 11
Db |||||
63 LSLRSAHLAQ 73

RESULT 15
AAR44024
ID AAR44024 standard; protein; 236 AA.
XX AC AAR44024;
XX XX
XX 25-MAR-2003 (revised)
DT 08-DEC-1993 (first entry)
XX XX
XX "Glu-114" E.coli heat labile toxin subunit A.
XX enterotoxigenic bacteria; vaccine; immunogenic detoxified LT-A;
KW protomer A; site-directed mutagenesis; reduced toxicity;
KW ADP-ribosyltransferase activity.
XX OS Escherichia coli.
XX XX
XX Key Location/Qualifiers
FH Misc-difference 110
FT /note= "corresponds to position 114 in cholera toxin A
FT subunit; wild-type Ser is substituted by Glu to reduce
FT toxicity"
XX XX
XX W09313202-A1.
XX XX
XX 08-JUL-1993.
XX XX
XX 30-DEC-1992; 92WO-EP003016.
XX XX
XX 31-DEC-1991; 91IT-MI003513.
XX XX
XX (BIOC-) BIOGINE SCLAVO SPA.
XX PI Domenighini M, Rappuoli R, Pizza M, Hol W;
XX XX
XX WPI; 1993-227320/28.
XX DR N-PSDB; AAQ51325.
XX XX
XX Immunogenic detoxified mutant cholera toxin and heat labile toxin -
PT

```

```

PT useful as vaccines against infection by Vibrio cholerae and enterotoxin
PT producing Escherichia coli.
XX XX
XX Claim 3; Fig 2 and Page 46; 60pp; English.
XX CC
XX CC The wild-type sequence coding for the A subunit of the heat labile toxin
CC (LT-A) of a strain of E.coli known to affect humans (Yamamoto et al,
CC J.Biol. Chem., 259, 5037-5044 - see AAQ42768) was subjected to site-
CC directed mutagenesis. Certain mutations were found to reduce toxicity
CC (see AAR38730-32 and AAR44016-R44025). The invention relates to
CC immunogenic, detoxified LT-A proteins and their use in vaccines to
CC protect against enterotoxigenic E.coli. The amino acid sequence of this
CC preferred detoxified mutain is not printed in the specification but has
CC been assembled from the full-length wild-type sequence and the
CC description given in the text. (N.B. Amino acid numbering is based on the
CC cholera toxin A subunit sequence). (Updated on 25-MAR-2003 to correct PN
XX field.) (Updated on 25-MAR-2003 to correct PR field.)
XX SQ Sequence 236 AA;

Query Match 68.8%; Score 11; DB 2; Length 236;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSAHLAQ 11
Db |||||
63 LSLRSAHLAQ 73

Search completed: April 23, 2004, 14:51:36
Job time : 49 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:49:06 ; Search time 14 Seconds
(without alignments)

59.001 Million cell updates/sec

Title: US-09-528-682-1_COPY_64_79

Perfect score: 16

Sequence: 1 LSLRSHLAGQSILSG 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/1aa/5A COMB pep:*
- 2: /cgn2_6/prodata/2/1aa/5B COMB pep:*
- 3: /cgn2_6/prodata/2/1aa/6A COMB pep:*
- 4: /cgn2_6/prodata/2/1aa/6B COMB pep:*
- 5: /cgn2_6/prodata/2/1aa/PCITUS COMB pep:*
- 6: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	16	100.0	240	3	US-08-823-120-3
2	11	68.8	236	3	US-08-823-120-6
3	11	68.8	254	3	US-08-486-099-114
4	11	68.8	254	3	US-08-360-107A-124
5	11	68.8	254	3	US-08-484-223B-114
6	11	68.8	254	3	US-08-919-597-114
7	11	68.8	254	3	US-08-475-668A-114
8	11	68.8	254	3	US-08-485-551A-114
9	11	68.8	254	3	US-08-471-913A-114
10	11	68.8	254	3	US-08-485-264A-114
11	11	68.8	254	4	US-08-474-349A-114
12	11	68.8	254	4	US-08-470-896-114
13	11	68.8	254	4	US-08-485-546A-114
14	7	43.8	12	2	US-08-292-968-7
15	7	43.8	12	2	US-08-467-974-7
16	7	43.8	12	2	US-08-467-536-7
17	7	43.8	12	2	US-08-467-976-7
18	7	43.8	12	3	US-09-082-514-7
19	7	43.8	20	1	US-08-171-299B-11
20	7	43.8	179	2	US-08-435-605A-57
21	7	43.8	192	2	US-08-435-605A-8
22	7	43.8	194	2	US-08-435-605A-6
23	7	43.8	194	2	US-08-435-605A-14
24	7	43.8	194	2	US-08-435-605A-16
25	7	43.8	194	2	US-08-435-605A-54
26	7	43.8	194	2	US-08-435-605A-55
27	7	43.8	215	4	US-09-107-532A-5710

Sequence 2, Appli
Sequence 13, Appl
Sequence 15, Appl
Sequence 49, Appl
Sequence 51, Appl
Sequence 52, Appl
Sequence 53, Appl
Sequence 56, Appl
Sequence 4, Appl
Sequence 8, Appl
Sequence 2, Appl
Sequence 2, Appl
Patent No. 5223610
Sequence 27252, A
Sequence 6175, Ap
Sequence 1413, Ap
Sequence 1432, Ap
Sequence 10, Appl

28 7 43.8 238 3 US-08-823-120-2
29 7 43.8 240 2 US-08-435-605A-13
30 7 43.8 240 2 US-08-435-605A-15
31 7 43.8 240 2 US-08-435-605A-49
32 7 43.8 240 2 US-08-435-605A-51
33 7 43.8 240 2 US-08-435-605A-52
34 7 43.8 240 2 US-08-435-605A-53
35 7 43.8 240 2 US-08-435-605A-56
36 7 43.8 240 3 US-08-823-120-4
37 7 43.8 240 3 US-08-823-120-8
38 7 43.8 258 1 US-08-449-045C-2
39 7 43.8 258 2 US-08-435-605A-2
40 7 43.8 258 6 5223610-2
41 7 43.8 412 4 US-09-252-991A-27252
42 7 43.8 631 4 US-09-134-000C-6175
43 6 37.5 7 4 US-09-989-789-1413
44 6 37.5 7 4 US-09-989-789-1432
45 6 37.5 20 1 US-08-171-299B-10

ALIGNMENTS

RESULT 1

US-08-823-120-3
; Sequence 3, Application US/08923120
; Patent No. 6149919

; GENERAL INFORMATION:
; APPLICANT: Domenighini, Mario
; APPLICANT: Rappuoli, Rino

; APPLICANT: Pizzia, Mariagrazia
; TITLE OF INVENTION: Immunogenic Detoxified Mutants of

; TITLE OF INVENTION: Cholera Toxin and of the Toxin Lt, Their Preparation and
; NUMBER OF SEQUENCES: 41
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESS: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,120

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/256,003
; FILING DATE: 11-NOV-1994
; ATTORNEY/AGENT INFORMATION:

; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0315.001

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-823-120-3

Query Match 100.0%; Score 16; DB 3; Length 240;

Best Local Similarity 100.0%; Pred. No. 6.7e-09;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLSRSAHLAQCSILSG 16
Db 64 LLSRSAHLAQCSILSG 79

RESULT 2

US-08-823-120-6
; Sequence 6, Application US/08823120
; Patent No. 6149919
; GENERAL INFORMATION:
; APPLICANT: Domenighini, Mario
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; TITLE OF INVENTION: Immunogenic Detoxified Mutants of
; TITLE OF INVENTION: Cholera Toxin and of the Toxin Lt, Their Preparation and
; TITLE OF INVENTION: Their Use for the Preparation of Vaccines
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,120
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,003
; FILING DATE: 11-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0315.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-823-120-6

Query Match 68.8%; Score 11; DB 3; Length 236;
Best Local Similarity 100.0%; Pred. No. 0.00067;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLSRSAHLAQ 11
Db 63 LLSRSAHLAQ 73

RESULT 3

US-08-486-099-114
; Sequence 114, Application US/08486099
; Patent No. 6013283
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.

; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HEPATITIS
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,099
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-486-099-114
Query Match 68.8%; Score 11; DB 3; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.00072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LLSRSAHLAQ 11
Db 81 LLSRSAHLAQ 91
RESULT 4
US-08-360-107A-124
; Sequence 124, Application US/08360107A
; Patent No. 6017536
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 149
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,107A
; FILING DATE: 20-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 124:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-360-107A-124

Query Match 68.8%; Score 11; DB 3; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.00072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSAHLAQ 11
|||/|||/|||
Db 81 LSLRSAHLAQ 91

RESULT 5
US-08-484-223B-114
; Sequence 114, Application US/08484223B
; Patent No. 6020459
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; NUMBER OF SEQUENCES: 245
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,223B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-484-223B-114

Query Match 68.8%; Score 11; DB 3; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.00072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSAHLAQ 11

; LENGTH: 254 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-484-223B-114

Query Match 68.8%; Score 11; DB 3; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.00072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSAHLAQ 11
|||/|||/|||
Db 81 LSLRSAHLAQ 91

RESULT 6
US-08-919-597-114
; Sequence 114, Application US/08919597
; Patent No. 6054265
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/470,896
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-919-597-114

Db 81 LSLRSALHAGQ 91
|||||

RESULT 7
US-08-475-668A-114
; Sequence 114, Application US/08475668A
; Patent No. 6060065
; GENERAL INFORMATION:
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 211
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,668A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 amino acids
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-475-668A-114

Query Match 68.8%; Score 11; DB 3; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.00072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LSLRSALHAGQ 11
Db 81 LSLRSALHAGQ 91
|||||

RESULT 8
US-08-485-551A-114
; Sequence 114, Application US/08485551A
; Patent No. 6068973
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
; TITLE OF INVENTION: TRANSMISSION

; NUMBER OF SEQUENCES: 211
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,551A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 amino acids
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-485-551A-114

Query Match 68.8%; Score 11; DB 3; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.00072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LSLRSALHAGQ 11
Db 81 LSLRSALHAGQ 91
|||||

RESULT 9
US-08-471-913A-114
; Sequence 114, Application US/08471913A
; Patent No. 6093794
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 214
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,913A

;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Coruzzi, Laura A.
;; REGISTRATION NUMBER: 30,742
;; REFERENCE/DOCKET NUMBER: 7872-030
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 790-9090
;; TELEFAX: (212) 869-9741/8864
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 114:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 254 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: unknown
;; MOLECULE TYPE: protein
US-08-471-913A-114
Query Match 68.8%; Score 11; DB 3; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.00072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LSLRSALHLAGQ 11
Db 81 LSLRSALHLAGQ 91
RESULT 10
US-08-485-264A-114
; Sequence 114, Application US/08485264A
; Patent No. 6228963
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING
; TITLE OF INVENTION: RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US/08/485,264A
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 amino acids
; TYPE: amino acid
; STRANDEDNESS:

;; TOPOLOGY: unknown
;; MOLECULE TYPE: protein
US-08-485-264A-114
Query Match 68.8%; Score 11; DB 3; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.00072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LSLRSALHLAGQ 11
Db 81 LSLRSALHLAGQ 91
RESULT 11
US-08-474-349A-114
; Sequence 114, Application US/08474349A
; Patent No. 6333395
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA
; TITLE OF INVENTION: VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 517
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US/08/474,349A
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
;; MOLECULE TYPE: protein
US-08-474-349A-114
Query Match 68.8%; Score 11; DB 4; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.00072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LSLRSALHLAGQ 11
Db 81 LSLRSALHLAGQ 91
RESULT 12
US-08-470-896-114


```
; Sequence 114, Application US/08470896
; Patent No. 6479055
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Pettaway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,896
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-470-896-114

Query Match 68.8%; Score 11; DB 4; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.00072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LSLRSALHAGQ 11
Db 81 LSLRSALHAGQ 91

RESULT 13
US-08-485-546A-114
; Sequence 114, Application US/08485546A
; Patent No. 6518013
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Pettaway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE
; FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 214
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSER: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,546A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-485-546A-114

Query Match 68.8%; Score 11; DB 4; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.00072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LSLRSALHAGQ 11
Db 81 LSLRSALHAGQ 91

RESULT 14
US-08-292-968-7
; Sequence 7, Application US/08292968
; Patent No. 5856122
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: COHEN, Raymond P.
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: KLEIN, Michel H.
; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/292,968
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```

; APPLICATION NUMBER: US 08/110,947
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/251,121
; FILING DATE: 31-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-388
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-292-968-7

```

```

Query Match 43.8%; Score 7; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 LSLRSAH 7
Db 6 LSLRSAH 12

```

RESULT 15

```

US-08-467-974-7
; Sequence 7, Application US/08467974
; Patent No. 5965385

```

GENERAL INFORMATION:

```

; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: KLEIN, Michel H.
; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,974
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/467,536
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 22-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/251,121
; FILING DATE: 31-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/110,947
; FILING DATE: 24-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.

```

```

; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-454 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-467-974-7

```

```

Query Match 43.8%; Score 7; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 LSLRSAH 7
Db 6 LSLRSAH 12

```

```

Search completed: April 23, 2004, 14:55:48
Job time : 14 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:54:17 ; Search time 35.6667 Seconds
(without alignments)
124.026 Million cell updates/sec

Title: US-09-528-682-1_COPY_64_79

Perfect score: 16
Sequence: 1 LSLRSALHAGQSILSG 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1133595 seqs, 276475211 residues

Word size : 0

Total number of hits satisfying chosen parameters: 11133595

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09C_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	100.0	240	9	US-09-819-917-7
2	16	100.0	240	10	US-09-297-171-1
3	16	100.0	240	14	US-10-304-496-1
4	16	100.0	259	14	US-10-334-729-4
5	11	68.8	254	12	US-10-267-682-114
6	11	68.8	254	12	US-10-267-748-114
7	7	43.8	236	9	US-09-819-917-6
8	7	43.8	240	9	US-09-819-917-8
9	7	43.8	258	14	US-10-334-729-46
10	7	43.8	258	14	US-10-334-729-48
11	7	43.8	258	14	US-10-334-729-50
12	7	43.8	258	14	US-10-334-729-52
13	7	43.8	342	12	US-10-381-882-2
14	7	43.8	362	9	US-09-809-033A-3
15	7	43.8	382	9	US-09-809-033A-4

16	7	43.8	556	10	US-09-949-029-36	Sequence 36, Appl
17	7	43.8	833	12	US-10-282-122A-68007	Sequence 68007, A
18	7	43.8	1209	12	US-10-381-882-6	Sequence 6, Appl
19	6	37.5	7	9	US-09-989-789-1413	Sequence 1413, Ap
20	6	37.5	7	9	US-09-989-789-1432	Sequence 1432, Ap
21	6	37.5	7	10	US-09-990-186-1413	Sequence 1413, Ap
22	6	37.5	7	10	US-09-990-186-1432	Sequence 1432, Ap
23	6	37.5	7	10	US-09-989-994-1413	Sequence 1413, Ap
24	6	37.5	7	10	US-09-989-994-1432	Sequence 1432, Ap
25	6	37.5	28	9	US-09-864-761-40492	Sequence 40492, A
26	6	37.5	38	14	US-10-097-111-500	Sequence 500, App
27	6	37.5	55	9	US-09-925-300-1478	Sequence 1478, Ap
28	6	37.5	81	12	US-10-424-599-223334	Sequence 223334,
29	6	37.5	84	11	US-09-864-408A-3918	Sequence 3918, Ap
30	6	37.5	84	14	US-10-029-386-33340	Sequence 33340, A
31	6	37.5	85	14	US-10-029-386-34025	Sequence 34025, A
32	6	37.5	87	12	US-10-424-599-160317	Sequence 160317,
33	6	37.5	107	11	US-09-864-408A-3052	Sequence 3052, Ap
34	6	37.5	109	15	US-10-104-047-2903	Sequence 2903, Ap
35	6	37.5	110	9	US-09-738-626-6835	Sequence 6835, Ap
36	6	37.5	124	12	US-10-424-599-227432	Sequence 227432,
37	6	37.5	134	9	US-09-864-761-46114	Sequence 46114, A
38	6	37.5	145	15	US-10-094-749-2679	Sequence 2679, Ap
39	6	37.5	156	12	US-10-425-114-72189	Sequence 72189, A
40	6	37.5	214	12	US-10-424-599-266355	Sequence 266355,
41	6	37.5	218	9	US-09-815-242-11644	Sequence 11644, A
42	6	37.5	279	12	US-10-424-599-210150	Sequence 210150,
43	6	37.5	282	14	US-10-156-761-7782	Sequence 7782, Ap
44	6	37.5	314	15	US-10-369-493-10341	Sequence 10341, A
45	6	37.5	327	16	US-10-389-566-507	Sequence 507, App

ALIGNMENTS

RESULT 1

US-09-819-917-7
; Sequence 7, Application US/09819917
; Patent No. US20020044939A1
; GENERAL INFORMATION:
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Fontana, Maria Rita
; APPLICANT: Giannelli, Valentina
; APPLICANT: Rappuoli, Rina
; TITLE OF INVENTION: Immunogenic Detoxified Mutants Of Cholera Toxin
; FILE REFERENCE: CHIR0312
; CURRENT APPLICATION NUMBER: US/09/819,917
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: 08/981,208
; PRIOR FILING DATE: 1997-12-22
; PRIOR APPLICATION NUMBER: 9513371.6
; PRIOR FILING DATE: 1995-06-30
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 7
; LENGTH: 240
; TYPE: PRT
; ORGANISM: E. coli
US-09-819-917-7

Query Match 100.0%; Score 16; DB 9; Length 240;
Best Local Similarity 100.0%; Pred. No. 4.4e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSALHAGQSILSG 16

Db 64 LSLRSALHAGQSILSG 79

RESULT 2

US-09-297-171-1
; Sequence 1, Application US/09297171
; Publication No. US20030113338A1

```
; GENERAL INFORMATION:
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Giuliani, Marzia M
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: IMMUNOGENIC DETOXIFIED MUTANT E. COLI LT-A TOXIN
; FILE REFERENCE: 0342.001 / 2302-0342
; CURRENT APPLICATION NUMBER: US/09/297,171
; CURRENT FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: PCT/IB97/01440
; PRIOR FILING DATE: 1997-10-30
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-297-171-1

Query Match      100.0%; Score 16; DB 10; Length 240;
Best Local Similarity 100.0%; Pred. No. 4.4e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSAHLAQSIILSG 16
   |||||
Db 64 LSLRSAHLAQSIILSG 79

RESULT 3
US-10-304-496-1
; Sequence 1, Application US/10304496
; Publication No. US20030170262A1
; GENERAL INFORMATION:
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Giuliani, Marzia M
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: IMMUNOGENIC DETOXIFIED MUTANT E. COLI LT-A TOXIN
; FILE REFERENCE: 0342.001 / 2302-0342
; CURRENT APPLICATION NUMBER: US/10/304,496
; CURRENT FILING DATE: 2002-11-25
; PRIOR APPLICATION NUMBER: US/09/297,171
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: PCT/IB97/01440
; PRIOR FILING DATE: 1997-10-30
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-304-496-1

Query Match      100.0%; Score 16; DB 14; Length 240;
Best Local Similarity 100.0%; Pred. No. 4.4e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSAHLAQSIILSG 16
   |||||
Db 64 LSLRSAHLAQSIILSG 79

RESULT 4
US-10-334-729-4
; Sequence 4, Application US/10334729
; Publication No. US20030176653A1
; GENERAL INFORMATION:
; APPLICANT: Mason, Hugh
; APPLICANT: Arntzen, Charles
; TITLE OF INVENTION: Orally Immunogenic Bacterial Enterotoxins Expressed In
; FILE REFERENCE: 3121/1064
; CURRENT APPLICATION NUMBER: US/10/334,729
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: 09/470,124
```

```
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: 60/113,507
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:E. coli
; OTHER INFORMATION: heat-labile toxin gene mutagenized to optimize
; OTHER INFORMATION: expression in plants.
US-10-334-729-4

Query Match      100.0%; Score 16; DB 14; Length 259;
Best Local Similarity 100.0%; Pred. No. 4.7e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSAHLAQSIILSG 16
   |||||
Db 83 LSLRSAHLAQSIILSG 98

RESULT 5
US-10-267-682-114
; Sequence 114, Application US/10267682
; Publication No. US2004003235A1
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TRANSMISSION
; NUMBER OF SEQUENCES: 239
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/267,682
; FILING DATE: 08-Oct-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,223A
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: unknown
```

; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 114:
US-10-267-748-114

Query Match 68.8%; Score 11; DB 12; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.0038;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSLAHLAGQ 11
|||
DB 81 LSLRSLAHLAGQ 91

RESULT 6

US-10-267-748-114
; Sequence 114, Application US/10267748
; Publication No. US20040052820A1
; GENERAL INFORMATION:

APPLICANT: Bolognesi, Dani P.
Matthews, Thomas J.
Wild, Carl T.
Barney, Shawn O.
Lambert, Dennis M.
Petteway, Stephen R.
Langlois, Alphonse J.

TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
TRANSMISSION

NUMBER OF SEQUENCES: 239

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/267,748

FILING DATE: 08-Oct-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/484,223A

FILING DATE: 07-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Cortuzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7872-029

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 114:

SEQUENCE CHARACTERISTICS:

LENGTH: 254 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: unknown

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 114:

US-10-267-748-114

Query Match 68.8%; Score 11; DB 12; Length 254;

Best Local Similarity 100.0%; Pred. No. 0.0038;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LSLRSLAHLAGQ 11
|||
DB 81 LSLRSLAHLAGQ 91

RESULT 7

US-09-819-917-6
; Sequence 6, Application US/09819917
; Patent No. US20020044939A1
; GENERAL INFORMATION:

APPLICANT: Pizza, Mariagrazia

APPLICANT: Fontana, Maria Rita

APPLICANT: Giannelli, Valentina

APPLICANT: Rappuoli, Rina

TITLE OF INVENTION: Immunogenic Detoxified Mutants Of Cholera Toxin

FILE REFERENCE: CHIR0312

CURRENT APPLICATION NUMBER: US/09/819,917

CURRENT FILING DATE: 2001-03-28

PRIOR APPLICATION NUMBER: 08/981,208

PRIOR FILING DATE: 1997-12-22

PRIOR APPLICATION NUMBER: 9513371.6

PRIOR FILING DATE: 1995-06-30

NUMBER OF SEQ ID NOS: 8

SOFTWARE: PatentIn version 3.0

SEQ ID NO 6

LENGTH: 236

TYPE: PRT

ORGANISM: E. coli

US-09-819-917-6

Query Match 43.8%; Score 7; DB 9; Length 236;

Best Local Similarity 100.0%; Pred. No. 31;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SLRSLAH 8
|||
DB 64 SLRSLAH 70

RESULT 8

US-09-819-917-8
; Sequence 8, Application US/09819917
; Patent No. US20020044939A1
; GENERAL INFORMATION:

APPLICANT: Pizza, Mariagrazia

APPLICANT: Fontana, Maria Rita

APPLICANT: Giannelli, Valentina

APPLICANT: Rappuoli, Rina

TITLE OF INVENTION: Immunogenic Detoxified Mutants Of Cholera Toxin

FILE REFERENCE: CHIR0312

CURRENT APPLICATION NUMBER: US/09/819,917

CURRENT FILING DATE: 2001-03-28

PRIOR APPLICATION NUMBER: 08/981,208

PRIOR FILING DATE: 1997-12-22

PRIOR APPLICATION NUMBER: 9513371.6

PRIOR FILING DATE: 1995-06-30

NUMBER OF SEQ ID NOS: 8

SOFTWARE: PatentIn version 3.0

SEQ ID NO 8

LENGTH: 240

TYPE: PRT

ORGANISM: vibrio cholerae

US-09-819-917-8

Query Match 43.8%; Score 7; DB 9; Length 240;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SLRSLAH 8
|||
DB 65 SLRSLAH 71

RESULT 9

US-10-334-729-46
; Sequence 46, Application US/10334729

; Publication No. US20030176653A1
; GENERAL INFORMATION:
; APPLICANT: Mason, Hugh
; APPLICANT: Arntzen, Charles
; TITLE OF INVENTION: Orally Immunogenic Bacterial Enterotoxins Expressed In
; TITLE OF INVENTION: Transgenic Plants
; FILE REFERENCE: 3121/1064
; CURRENT APPLICATION NUMBER: US/10/334,729
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: 09/470,124
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: 60/113,507
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 46
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:V. cholerae
; OTHER INFORMATION: cholera toxin gene mutagenized to optimize
; OTHER INFORMATION: expression in plants.
US-10-334-729-46

Query Match 43.8%; Score 7; DB 14; Length 258;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SLRSAHL 8
| | | | |
Db 83 SLRSAHL 89

RESULT 10
US-10-334-729-48
; Sequence 48, Application US/10334729
; Publication No. US20030176653A1
; GENERAL INFORMATION:
; APPLICANT: Mason, Hugh
; APPLICANT: Arntzen, Charles
; TITLE OF INVENTION: Orally Immunogenic Bacterial Enterotoxins Expressed In
; TITLE OF INVENTION: Transgenic Plants
; FILE REFERENCE: 3121/1064
; CURRENT APPLICATION NUMBER: US/10/334,729
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: 09/470,124
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: 60/113,507
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 48
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:V. cholerae
; OTHER INFORMATION: cholera toxin gene mutagenized to optimize
; OTHER INFORMATION: expression in plants.
US-10-334-729-48

Query Match 43.8%; Score 7; DB 14; Length 258;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SLRSAHL 8
| | | | |
Db 83 SLRSAHL 89

RESULT 11
US-10-334-729-50

; Sequence 50, Application US/10334729
; Publication No. US20030176653A1
; GENERAL INFORMATION:
; APPLICANT: Mason, Hugh
; APPLICANT: Arntzen, Charles
; TITLE OF INVENTION: Orally Immunogenic Bacterial Enterotoxins Expressed In
; TITLE OF INVENTION: Transgenic Plants
; FILE REFERENCE: 3121/1064
; CURRENT APPLICATION NUMBER: US/10/334,729
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: 09/470,124
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: 60/113,507
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 50
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:V. cholerae
; OTHER INFORMATION: cholera toxin gene mutagenized to optimize
; OTHER INFORMATION: expression in plants.
US-10-334-729-50

Query Match 43.8%; Score 7; DB 14; Length 258;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SLRSAHL 8
| | | | |
Db 83 SLRSAHL 89

RESULT 12
US-10-334-729-52
; Sequence 52, Application US/10334729
; Publication No. US20030176653A1
; GENERAL INFORMATION:
; APPLICANT: Mason, Hugh
; APPLICANT: Arntzen, Charles
; TITLE OF INVENTION: Orally Immunogenic Bacterial Enterotoxins Expressed In
; TITLE OF INVENTION: Transgenic Plants
; FILE REFERENCE: 3121/1064
; CURRENT APPLICATION NUMBER: US/10/334,729
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: 09/470,124
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: 60/113,507
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 52
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:V. cholerae
; OTHER INFORMATION: cholera toxin gene mutagenized to optimize
; OTHER INFORMATION: expression in plants.
US-10-334-729-52

Query Match 43.8%; Score 7; DB 14; Length 258;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SLRSAHL 8
| | | | |
Db 83 SLRSAHL 89

RESULT 13

```
US-10-381-882-2
; Sequence 2, Application US/10381882
; Publication No. US20040052815A1
; GENERAL INFORMATION:
; APPLICANT: DYCKE, NILS
; APPLICANT: DALSGAARD, KRISTIAN
; APPLICANT: MC MOWAT, ALLAN
; APPLICANT: LOWENADLER, BJORN
; APPLICANT: KAASTRUP, PETER
; TITLE OF INVENTION: NEW IMMUNOGENIC COMPLEX
; FILE REFERENCE: 1501-1188
; CURRENT APPLICATION NUMBER: US/10/381,882
; PRIOR FILING DATE: 2003-08-19
; PRIOR APPLICATION NUMBER: PCT/SE01/02117
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: SE 0003538-6
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: CTAL-DD fusion protein
US-10-381-882-2

Query Match      43.8%; Score 7; DB 12; Length 342;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 SLRSAHL 8
      |||||
Db      75 SLRSAHL 81

RESULT 14
US-09-809-033A-3
; Sequence 3, Application US/09809033A
; Patent No. US20020142006A1
; GENERAL INFORMATION:
; APPLICANT: McGhee, Jerry
;             Kiyono, Hiroshi
;             Takeda, Yoshifumi
;             Ohmura, Mari
;             Yamamoto, Shingo
; TITLE OF INVENTION: Chimeric No. US20020142006A1toxic Mutants of
;                   Enterotoxins as Mucosal Adjuvants for Cell-Mediated or
;                   Humoral Immunity
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hendricks and Assoc
; STREET: P.O. Box 2509
; CITY: Fairfax
; STATE: VA
; COUNTRY: US
; ZIP: 22031
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/809,033A
; FILING DATE: 16-Mar-2001
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hendricks, Glenna
; REGISTRATION NUMBER: 32,535
; REFERENCE/DOCKET NUMBER: MCG-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703/425-8406
; TELEFAX: 703/425-8406
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 382 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-809-033A-4
Query Match      43.8%; Score 7; DB 9; Length 382;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 SLRSAHL 8
      |||||
Db      83 SLRSAHL 89

RESULT 15
US-09-809-033A-4
; Sequence 4, Application US/09809033A
; Patent No. US20020142006A1
; GENERAL INFORMATION:
; APPLICANT: McGhee, Jerry
;             Kiyono, Hiroshi
;             Takeda, Yoshifumi
;             Ohmura, Mari
;             Yamamoto, Shingo
; TITLE OF INVENTION: Chimeric No. US20020142006A1toxic Mutants of
;                   Enterotoxins as Mucosal Adjuvants for Cell-Mediated or
;                   Humoral Immunity
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hendricks and Assoc
; STREET: P.O. Box 2509
; CITY: Fairfax
; STATE: VA
; COUNTRY: US
; ZIP: 22031
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/809,033A
; FILING DATE: 16-Mar-2001
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hendricks, Glenna
; REGISTRATION NUMBER: 32,535
; REFERENCE/DOCKET NUMBER: MCG-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703/425-8405
; TELEFAX: 703/425-8406
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 382 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-809-033A-4
Query Match      43.8%; Score 7; DB 9; Length 382;
```

Best Local Similarity 100.0%; Pred. NO. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SLRSAHL 8
|||||
Db 83 SLRSAHL 89

Search completed: April 23, 2004, 15:07:24
Job time : 36.6667 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:48:31 ; Search time 11.6667 Seconds
(without alignments)
131.920 Million cell updates/sec

Title: US-09-528-682-3_COPY_64_79

Perfect score: 16

Sequence: 1 MSTFEQVNNKEFKG 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR 78: *

1: Pirl: *

2: Pirl: *

3: Pirl: *

4: Pirl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	9	56.2	259	2 A29831	heat-labile entero
2	6	37.5	295	2 E95200	fructokinase [impo
3	6	37.5	295	2 D98067	fructokinase [EC 2
4	6	37.5	296	2 D97052	homoserine kinase
5	6	37.5	296	2 AH1173	conserved hypothet
6	6	37.5	309	2 H95191	hypothetical prote
7	6	37.5	309	2 E98058	hypothetical prote
8	6	37.5	470	2 D69394	phosphoribosylamin
9	6	37.5	472	1 A35327	vitamin D-binding
10	6	37.5	476	1 VYR7D	vitamin D-binding
11	6	37.5	478	1 S61986	subtilisin-like pr
12	6	37.5	499	2 G82923	multiple banded an
13	6	37.5	531	2 E82295	translation releas
14	6	37.5	565	2 E70192	ABC transporter, A
15	6	37.5	663	1 A69798	beta-galactosidase
16	6	37.5	701	2 S51403	probable membrane
17	6	37.5	776	2 A48480	outer capsid prote
18	6	37.5	971	2 C82880	conserved hypothet
19	5	31.2	25	2 A61457	alpha-glucosidase
20	5	31.2	58	2 G95132	hypothetical prote
21	5	31.2	62	2 F69173	hypothetical prote
22	5	31.2	64	2 F70224	hypothetical prote
23	5	31.2	79	2 G64375	hypothetical prote
24	5	31.2	85	2 AB0351	hypothetical prote
25	5	31.2	87	2 T39465	probable acyl-coen
26	5	31.2	100	2 AD2566	hypothetical prote
27	5	31.2	107	2 S76457	hypothetical prote
28	5	31.2	114	2 T03016	probable apoptotic
29	5	31.2	115	2 S71269	apoptotic cell dea

30 5 31.2 115 2 F86446 probable defender
31 5 31.2 119 2 T17016 defender against c
32 5 31.2 123 1 R5ZM14 ribosomal protein
33 5 31.2 128 2 S32936 sporulation initia
34 5 31.2 136 2 I40774 hypothetical prote
35 5 31.2 136 2 D81361 hypothetical prote
36 5 31.2 145 2 JC2575 hypothetical 17K p
37 5 31.2 145 2 F81272 hypothetical prote
38 5 31.2 148 2 C90269 hypothetical prote
39 5 31.2 148 2 G84466 hypothetical prote
40 5 31.2 153 2 S77187 superoxide dismuta
41 5 31.2 154 2 S24971 probable 30S ribos
42 5 31.2 154 2 D96807 conserved hypothet
43 5 31.2 154 2 D87667 conserved hypothet
44 5 31.2 155 2 S76220 hypothetical prote
45 5 31.2 160 2 F84769 defender against c

ALIGNMENTS

RESULT 1

A29831 heat-labile enterotoxin IIA chain A precursor - Escherichia coli
N:Alternate names: LT-IIa
C:Species: Escherichia coli
C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: A29831
R:Pickett, C.L.; Weinstein, D.L.; Holmes, R.K.
J. Bacteriol. 169, 5180-5187, 1987
A:Title: Genetics of type IIA heat-labile enterotoxin of Escherichia coli: operon fusion
A:Reference number: A91849; MUID:88032841; PMID:2822667
A:Accession: A29831
A:Molecule type: DNA
A:Residues: 1-259 <PIC>
A:Cross-references: GB:M17894; NID:G146671; PIDN:AAA24093.1; PID:G146672
A>Note: the authors translated the codon TAT for residue 225 as Thr
C:Superfamily: heat-labile enterotoxin chain A
F:1-19/Domain: signal sequence #status predicted <SIG>
F:19-259/Product: heat-labile enterotoxin IIA chain A #status predicted <LTA>

Query Match 56.2%; Score 9; DB 2; Length 259;
Best Local Similarity 100.0%; Pred. No. 0.0055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 VPKNKEFKG 15
DB 204 VPKNKEFKG 212

RESULT 2

E95200 fructokinase [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001

C:Accession: E95200

R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: E95200

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-295 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK75798.1; PID:G14973216; GSPDB:GN00164; TIGR:SP4.

A:Experimental source: strain TIGR4

C:Genetics:

C:Superfamily: fructokinase; glucose kinase homology

Query Match 37.5%; Score 6; DB 2; Length 295;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFGKV 16
 |||||
 Db 171 KEFGKV 176

RESULT 3
 D98067
 fructokinase (EC 2.7.1.4) [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
 C:Accession: D98067
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; PMID:21429245; PMID:11544234
 A:Accession: D98067
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-295 <KUR>
 A:Cross-references: GB:AE007317; PIDN:AAL00369.1; PID:gl5459231; GSPDB:GN00174
 C:Genetics:
 A:Gene: scrK
 C:Superfamily: fructokinase; glucose kinase homology
 C:Keywords: phosphotransferase

Query Match 37.5%; Score 6; DB 2; Length 295;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFGKV 16
 |||||
 Db 171 KEFGKV 176

RESULT 4
 D97052
 homoserine kinase (thrB) [imported] - Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 12-May-2003
 C:Accession: D97052
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
 J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
 A:Reference number: A96900; PMID:21359325; PMID:21359325
 A:Accession: D97052
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-296 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK79207.1; PID:gl5024160; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC1235
 C:Superfamily: homoserine kinase

Query Match 37.5%; Score 6; DB 2; Length 296;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 NKEFGK 15
 |||||
 Db 69 NKEFGK 74

RESULT 5
 AH1173
 conserved hypothetical protein lmo0792 [imported] - Listeria monocytogenes (stra
 C:Species: Listeria monocytogenes
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AH1173
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kreft, J.; Kuhn, M.; Kunet, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlund,
 A:Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; PMID:21537279; PMID:11679669
 A:Accession: AH1173
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-296 <GLA>
 A:Cross-references: GB:NC_003210; PIDN:CAC98870.1; PID:gl6410181; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lmo0792

Query Match 37.5%; Score 6; DB 2; Length 296;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 VPNNKE 12
 |||||
 Db 14 VPNNKE 19

RESULT 6
 H95191
 hypothetical protein SPI650 [imported] - Streptococcus pneumoniae (strain TIGR4)
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
 C:Accession: H95191
 R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heic
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzap
 nson, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
 A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; PMID:21357209; PMID:11463916
 A:Accession: H95191
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-309 <KUR>
 A:Cross-references: GB:AE005672; PIDN:AAK75729.1; PID:gl4973140; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SPI650
 C:Superfamily: adhesin B

Query Match 37.5%; Score 6; DB 2; Length 309;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PNNKEF 13
 |||||
 Db 163 PNNKEF 168

RESULT 7
 E98058
 hypothetical protein psaA [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
 C:Accession: E98058
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: E98058

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-309 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAL00298.1; PID:g15459154; GSPDB:GN00174

C:Genetics:

A:Gene: psaA

C:Superfamily: adhesin B

Query Match 37.5%; Score 6; DB 2; Length 309;

Best Local Similarity 100.0%; Pred. No. 16;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PNNKEF 13

Db 163 PNNKEF 168

RESULT 8

D69394

phosphoribosylamine-glycine ligase (purD) homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000

C:Accession: D69394

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

R: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.

G: Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Woese, C.R.; Venter, J.C.

A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: D69394

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-470 <KLE>

A:Cross-references: GB:AE001024; GB:AE000782; NID:92689347; PIDN:AAB90089.1; PID:g264943

C:Superfamily: phosphoribosylamine-glycine ligase; phosphoribosylamine-glycine ligase hc

F:2-467/Domain: phosphoribosylamine-glycine ligase homology <PGL>

Query Match 37.5%; Score 6; DB 2; Length 470;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16

Db 299 KEFKGV 304

RESULT 9

A35327

vitamin D-binding protein precursor - mouse (fragment)

N:Alternate names: DBP; Gc-globulin; group-specific component

C:Species: Mus musculus (house mouse)

C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999

C:Accession: A35327

R:Yang, F.; Bergeron, J.M.; Linehan, L.A.; Lailey, P.A.; Sakaguchi, A.Y.; Bowman, B.H.

Genomics 7, 509-516, 1990

A>Title: Mapping and conservation of the group-specific component gene in mouse.

A:Reference number: A35327; MUID:90353947; PMID:1696927

A:Accession: A35327

A:Molecule type: mRNA

A:Residues: 1-472 <YAN>

A:Cross-references: GB:M55413; GB:J04762; NID:g193445; PIDN:AAA37669.1; PID:g193446

C:Comment: DBP is a multifunctional protein found in plasma, ascitic fluid, cerebrospine

nts polymerization of actin by binding its monomers. DBP associates with membrane-bound

C:Genetics:

A:Gene: GC

A:Map position: 5

C:Superfamily: serum albumin; serum albumin repeat homology

C:Keywords: actin binding; duplication; globulin; glycoprotein; liver; plasma; polymorph

F:1-12/Domain: signal sequence (fragment) #status predicted <SIG>

F:13-472/Product: vitamin D-binding protein #status predicted <MAT>

F:22-195/Domain: serum albumin repeat homology <SA1>

F:213-381/Domain: serum albumin repeat homology <SA2>

F:400-472/Domain: serum albumin repeat homology #status atypical <SA3>

F:25-71,70-79,92-108,107-118,141-186,185-194,216-262,261-269,282-296,295-307,331-372,371

F:284/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 37.5%; Score 6; DB 1; Length 472;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 STEQOV 7

Db 51 STEQOV 56

RESULT 10

VYRTD

vitamin D-binding protein precursor - rat

N:Alternate names: DBP; Gc-globulin; group-specific component

C:Species: Rattus norvegicus (Norway rat)

C>Date: 04-Dec-1986 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999

C:Accession: A38726; A34161; A92576; A93050; A03238

R:Ray, K.; Wang, X.; Zhao, M.; Cooke, N.E.

J. Biol. Chem. 266, 6221-6229, 1991

A>Title: The rat vitamin D binding protein (Gc-globulin) gene. Structural analysis, func

A:Reference number: A38726; MUID:91177870; PMID:2007578

A:Accession: A38726

A:Molecule type: DNA

A:Residues: 1-476 <RAY>

A:Cross-references: GB:M60197

A:Experimental source: liver

A>Note: the authors translated the codon CAG for residue 129 as Gly, CTT for residue 174

R:McLeod, J.F.; Cooke, N.E.

J. Biol. Chem. 264, 21760-21769, 1989

A>Title: The vitamin D-binding protein, alpha-fetoprotein, albumin multigene family: det

A:Reference number: A34161; MUID:90094352; PMID:2480956

A:Accession: A34161

A:Molecule type: mRNA

A:Residues: 1-131, 'Q', 133-476 <MCL>

A:Cross-references: GB:J05148; NID:g203940; PIDN:AAA41082.1; PID:g203941

A:Experimental source: kidney

R:Cooke, N.E.

J. Biol. Chem. 261, 3441-3450, 1986

A>Title: Rat vitamin D binding protein. Determination of the full-length primary structu

A:Reference number: A92576; MUID:86140127; PMID:2419332

A:Accession: A92576

A:Molecule type: mRNA

A:Residues: 1-173, 'P', 175-209, 'L', 211-476 <COO>

A:Cross-references: GB:M2450; NID:g203926; PIDN:AAA41080.1; PID:g203927

A:Experimental source: liver

R:Litwiler, R.; Fass, D.; Kumar, R.

Life Sci. 38, 2179-2184, 1986

A>Title: The amino acid sequence of the NH-2-terminal portion of rat and human vitamin D

A:Reference number: A93050; MUID:86229807; PMID:3713442

A:Accession: A93050

A:Molecule type: protein

A:Residues: 17-40 <LIT>

C:Comment: DBP is a multifunctional protein found in plasma, ascitic fluid, cerebrospina

nts polymerization of actin by binding its monomers. DBP associates with membrane-bound

C:Genetics:

A:Introns: 20/1, 43/2, 87/3, 158/2, 202/3, 234/2, 277/3, 345/2, 388/3; 421/2; 465/3

C:Superfamily: serum albumin; serum albumin repeat homology

C:Keywords: actin binding; duplication; globulin; glycoprotein; liver; plasma; polymorph

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-476/Product: vitamin D-binding protein #status experimental <MP>

F:26-199/Domain: serum albumin repeat homology <SA1>

F:217-385/Domain: serum albumin repeat homology <SA2>

F:404-476/Domain: serum albumin repeat homology #status atypical <SA3>

F:29-75,74-83,96-112,111-122,145-190,189-198,220-266,265-273,286-300,299-311,335-376,375

F:288/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 37.5%; Score 6; DB 1; Length 476;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 STFEQV 7
DB 55 STFEQV 60

RESULT 11
S61986
subtilisin-like proteinase (EC 3.4.21.-) YSP3 precursor - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein O2517; protein UNC478; protein YOR003w; subtilisin-like prote
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: S61986; S66868; S72135
R;Sterky, F.; Uhlen, M.
submitted to the EMBL Data Library, December 1995
A;Reference number: S61981
A;Accession: S61986
A;Molecule type: DNA
A;Residues: 1-478 <STW>
A;Cross-references: EMBL:U43491; NID:g1150992; PIDN:AAC49482.1; PID:g1150998
R;Pettersson, B.; Sterky, F.; Uhlen, M.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S66868
A;Accession: S66868
A;Molecule type: DNA
A;Residues: 1-478 <PET>
A;Cross-references: EMBL:Z74911; NID:g1420091; PIDN:CAA99191.1; PID:g1420092; GSPDB:GN00
A;Experimental source: strain S288C
R;Sterky, F.; Holmberg, A.; Pettersson, B.; Uhlen, M.
Yeast 12, 1091-1095, 1996
A;Title: The sequence of a 30 kb fragment on the left arm of chromosome XV from Sacchar
A;Reference number: S72130; MUID:97051599; PMID:8896276
A;Accession: S72135
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-478 <STW>
A;Cross-references: EMBL:U43491; NID:g1150992; PIDN:AAC49482.1; PID:g1150998
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1995
C;Genetics:
A;Gene: SGD:YSP3; MIPS:YOR003W
A;Cross-references: SGD:S0005529; MIPS:YOR003W
A;Map position: 15R
C;Superfamily: subtilisin; subtilisin homology
C;Keywords: hydrolase; serine proteinase
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-478/Product: subtilisin-like proteinase YSP3 #status predicted <MAT>
F;204-421/Domain: subtilisin homology <SST>
F;213,245,407/Active site: Asp, His, Ser #status predicted

Query Match 37.5%; Score 6; DB 1; Length 478;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 NKSPFK 15
DB 302 NKSPFK 307

RESULT 12
G82923
multiple banded antigen homolog UUI72 [imported] - Ureaplasma urealyticum
C;Species: Ureaplasma urealyticum
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: G82923
R;Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A;Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
A;Reference number: G82920
A;Accession: G82923
A;Status: preliminary

A;Molecule type: DNA
A;Residues: 1-499 <GLA>
A;Cross-references: GB:AE002117; GB:AF222894; NID:G6899132; PIDN:AAF30579.1; GSPDB:GN001
A;Experimental source: serovar 3; biovar 1
C;Genetics:
A;Gene: UUI72
A;Genetic code: SGC3

Query Match 37.5%; Score 6; DB 2; Length 499;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TPEQVP 8
DB 144 TPEQVP 149

RESULT 13
E82295
translation releasing factor RF-3 VC0659 [imported] - Vibrio cholerae (strain N16961 ser
N;Alternate names: peptide chain release factor 3
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: E82295
R;Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Emolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F
I, R.K.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: E82295
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-531 <HEI>
A;Cross-references: GB:AE004152; GB:AE003852; NID:G9655096; PIDN:AAF93825.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC0659
A;Map position: 1

Query Match 37.5%; Score 6; DB 2; Length 531;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
DB 179 KEFKGV 184

RESULT 14
E70192
ABC transporter, ATP-binding protein homolog - Lyme disease spirochete
C;Species: Borrelia burgdorferi (lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 02-Feb-2001
C;Accession: E70192
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: E70192
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-565 <KLE>
A;Cross-references: GB:AE001174; GB:AE000783; NID:g2688675; PIDN:AAC67091.1; PID:g268868
A;Experimental source: strain B31
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C;Keywords: ATP; nucleotide binding; P-loop
F;39-250/Domain: ATP-binding cassette homology <ABC1>
F;56-63/Region: nucleotide-binding motif A (P-loop)
F;357-531/Domain: ATP-binding cassette homology <ABC2>

```

Query Match          37.5%; Score 6; DB 2; Length 565;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      11 KEFKGV 16
      |||||
Db      502 KEFKGV 507

RESULT 15
A69798
beta-galactosidase homolog yesZ - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: A69798
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.: Bron, S.; Brouillet, S.; Bruchli, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: A69798
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-663 <KUN>
A:Cross-references: GB:Z99107; GB:AL009126; NID:g2632866; PIDN:CAB12527.1; PID:g2633021
A:Experimental source: strain 168
C:Genetics:
A:Gene: yesZ
C:Superfamily: Bacillus beta-galactosidase

Query Match          37.5%; Score 6; DB 1; Length 663;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 TFEQVP 8
      |||||
Db      191 TFEQVP 196

Search completed: April 23, 2004, 14:54:56
Job time : 13.6667 secs

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:47:06 ; Search time 8 Seconds
(without alignments)
104.140 Million cell updates/sec

Title: US-09-528-682-3_COPY_64_79
Perfect score: 16
Sequence: 1 MSTFEQVNNKEFGV 16

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Match	Length	ID	Description
1	9	56.2	259	1 E2AA_ECOLI	P13810 escherichia
2	6	37.5	296	1 KHSE_CLOAB	Q971n8 clostridium
3	6	37.5	309	1 MTSI_STRPN	P72538 streptococc
4	6	37.5	309	1 MTSA_STREMT	Q915X0 streptococc
5	6	37.5	464	1 CAP_DICDI	P54654 dictyosteli
6	6	37.5	470	1 PUR2_ARCFU	O29108 archaeoglob
7	6	37.5	476	1 VTDB_MOUSE	P21614 mus musculu
8	6	37.5	476	1 VTDB_RAT	P04276 rattus norv
9	6	37.5	478	1 YSP3_YEAST	P25036 saccharomyc
10	6	37.5	513	1 VNNI_HUMAN	Q95497 homo sapien
11	6	37.5	529	1 RF3_VIBPA	Q87m18 vibrio para
12	6	37.5	529	1 RF3_VIBVU	Q8dbt6 vibrio vuln
13	6	37.5	529	1 RF3_VIBVY	Q7m334 vibrio vuln
14	6	37.5	531	1 RF3_VIBCH	Q9ku64 vibrio chol
15	6	37.5	701	1 YL66_YEAST	Q06149 saccharomyc
16	6	37.5	1087	1 DP2L_THEAC	Q9hm33 thermoplas
17	6	37.5	1088	1 DP2L_THEVO	Q97cr6 thermoplas
18	5	31.2	79	1 DADI_MAIZE	Q81214 zea mays (m
19	5	31.2	79	1 Y607_METJA	Q58024 methanococc
20	5	31.2	114	1 DADI_HORVU	Q9sme9 hordeum vul
21	5	31.2	114	1 DADI_ORYSA	O50070 oryza sativ
22	5	31.2	114	1 DADI_HORVU	Q9sme8 hordeum vul
23	5	31.2	115	1 DADI_ARATH	Q93080 arabidopsis
24	5	31.2	115	1 DADI_BETVE	Q9m3t9 betula verr
25	5	31.2	115	1 DADI_CITUN	Q9zww7 citrus unsh
26	5	31.2	115	1 DADI_ARATH	O22622 arabidopsis
27	5	31.2	116	1 DADI_LYCES	Q9smc4 lycopersico
28	5	31.2	117	1 DADI_PEA	Q9zra3 pisum sativ
29	5	31.2	119	1 DADI_MALDO	O24060 malus domes
30	5	31.2	123	1 RK14_MAIZE	P08529 zea mays (m
31	5	31.2	128	1 KAPE_BACSU	Q08429 bacillus su
32	5	31.2	131	1 YH24_PYRFU	Q8u077 pyrococcus
33	5	31.2	148	1 YD36_METJA	Q58732 methanococc

34	5	31.2	154	1 RT13_ARATH	Q9cal9 arabidopsis
35	5	31.2	154	1 SODC_SCHPO	P28758 schizosacch
36	5	31.2	165	1 RL12_YEAST	P17079 saccharomyc
37	5	31.2	182	1 PMFF_PROMI	P53521 proteus mir
38	5	31.2	188	1 EFP_VIBCH	Q9kne1 vibrio chol
39	5	31.2	188	1 EFP_XANAC	Q8pjz7 xanthomonas
40	5	31.2	188	1 EFP_XYLEFA	Q9pam3 xylella fas
41	5	31.2	191	1 Y486_BUCBP	Q89a55 buchnera ap
42	5	31.2	196	1 PTH_VIBCH	Q9kqz1 vibrio chol
43	5	31.2	196	1 PTH_VIBVU	Q8df44 rhizomucor
44	5	31.2	198	1 RAS2_RHIRA	P22279 rhizomucor
45	5	31.2	198	1 YF77_PVRAB	Q9uydl pyrococcus

ALIGNMENTS

RESULT 1
E2AA_ECOLI STANDARD; PRT; 259 AA.
AC P13810;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Heat-labile enterotoxin IIA, A chain precursor (LT-IIA).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
ON NCBI_TaxID=562;
RX MEDLINE=8032841; PubMed=2822667;
RX Fickett C.L., Weinstein D.L., Holmes R.K.;
RT "Genetics of type IIA heat-labile enterotoxin of Escherichia coli:
RT operon fusions, nucleotide sequence, and hybridization studies.";
RL J. Bacteriol. 169:5180-5187(1987).
CC -!- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC -!- SUBUNIT: HETEROXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; M17894; AAA24093.1; -.
DR PIR; A29831; A29831.
DR HSSP; P43528; LTII.
DR InterPro; IPR001144; Enterotoxin_A.
DR Pfam; PF01375; Enterotoxin_A; 1.
DR PRINTS; PR00771; ENTEROTOXINA.
KW Enterotoxin; signal.
FT SIGNAL 1 18
FT CHAIN 19 259 HEAT-LABILE ENTEROTOXIN IIA, A CHAIN.
FT DISULFID 203 215 BY SIMILARITY.
FT ACT_SITE 128 128 BY SIMILARITY.
SQ SEQUENCE 259 AA; 29242 MW; 996F311A32CABEAA CRC64;

Query Match 56.2%; Score 9; DB 1; Length 259;
Best Local Similarity 100.0%; Pred.No. 0.0023;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QV 7 VPNNKEFGK 15
|||
Db 204 VPNNKEFGK 212

RESULT 2
KHSE_CLOAB STANDARD; PRT; 296 AA.
ID KHSE_CLOAB

AC O97JN8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Homoserine kinase (EC 2.7.1.39) (HK).
 GN THRB OR CAC1235.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OC NCBI_TaxID=1488;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Berton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gleason R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum";
 RL J. Bacteriol. 183:4823-4838(2001).
 CC -1- CATALYTIC ACTIVITY: ATP + L-homoserine = ADP + O-phospho-L-
 CC homoserine.
 CC -1- PATHWAY: Threonine biosynthesis from aspartate; fourth step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- SIMILARITY: Belongs to the GMP kinase family. Homoserine kinase
 CC subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AE007636; AAK79207.1; --
 CC PIR: D97052; D97052.
 CC HAMAP: MF_00384; --; 1.
 CC InterPro: IPR006204; GMP kinase.
 CC InterPro: IPR006203; GMPKase ATP.
 CC InterPro: IPR000870; Homoser Kin.
 CC Pfam: PF00288; GMP kinases; 1.
 CC PRINTS: PR00958; HOMSERKINASE.
 CC TIGRFAMs: TIGR00191; thrB; 1.
 CC PROSITE: PS00627; GMP_KINASES ATP; 1.
 CC Threonine biosynthesis; Transferase; Kinase; ATP-binding;
 CC Complete proteome.
 CC NP BIND 85 95 ATP (POTENTIAL).
 CC SEQUENCE 296 AA; 33072 MW; 809830C84B851346 CRC64;
 Query Match 37.5%; Score 6; DB 1; Length 296;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 10 NKEFGK 15
 Db 69 NKEFGK 74
 |||||
 RESULT 3
 ID MTAL STRPN STANDARD; PRT: 309 AA.
 AC P72538; Q54720; Q9L5X2; Q9L5X3; Q9L5X4; Q9R6P5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Manganese ABC transporter substrate-binding lipoprotein precursor
 DE (Pneumococcal surface adhesin A).
 GN PSAA OR SP1650 OR SP1494.
 OS Streptococcus pneumoniae, and
 OC Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.
 OX NCBI_TaxID=1313, 171101;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=6B;
 RA Sampson J.S., Whitney A.M., Furlow Z.;
 RT "Streptococcus pneumoniae surface adhesin A";
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=D39 / NCTC 7466 / Serotype 2;
 RX MEDLINE=97101047; PubMed=8945574;
 RA Berty A.M., Paton J.C.;
 RT "Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin
 RT essential for virulence of Streptococcus pneumoniae";
 RL Infect. Immun. 64:5255-5262(1996).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=98449534; PubMed=9767595;
 RA Novak R., Braun J.S., Charpentier E., Tuomanen E.;
 RT "Penicillin tolerance genes of Streptococcus pneumoniae: the ABC-type
 RT manganese permease complex Psa";
 RL Mol. Microbiol. 29:1285-1296(1998).
 RN [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=NA-1064/97, NA-1283/96, NA-1383/97, NA-1508/92, and
 RC ATCC BAA-255 / R6;
 RA Perez A., Jado I., Casal J.;
 RT "Identification of a psaA gene in viridans streptococcal strains";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC BAA-334 / TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae";
 RL Science 293:498-506(2001).
 RN [6]
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC BAA-255 / R6;
 RX MEDLINE=21429245; PubMed=11544234;
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
 RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
 RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
 RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
 RA McAhern S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
 RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaakunas S.R., Rostock P.R. Jr., Skatrud P.L.,
 RA Glass J.L.;
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6";
 RL J. Bacteriol. 183:5709-5717(2001).
 RN [7]
 RN FUNCTION.
 RP MEDLINE=98025470; PubMed=9379902;
 RA Dintilhac A., Allouing G., Granadel C., Claverys J.-P.;
 RT "Competence and virulence of Streptococcus pneumoniae: Adc and PsaA
 RT mutants exhibit a requirement for Zn and Mn resulting from
 RT inactivation of putative ABC metal permeases";
 RL Mol. Microbiol. 25:727-739(1997).
 CC -1- FUNCTION: Part of an ATP-driven transport system for manganese.
 CC Also act as an adhesin which is involved on adherence to
 CC extracellular matrix. It is an important factor in pathogenesis
 CC and infection.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor

(By similarity).
 -!- SIMILARITY: Belongs to the bacterial solute-binding protein family
 9. Lipoprotein receptor antigen (Lral) subfamily.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U53509; AAB09440.1; --
 DR EMBL; U40786; AAC24470.1; --
 DR EMBL; AF055088; AAD09975.1; --
 DR EMBL; AF248229; AAF70663.1; --
 DR EMBL; AF248230; AAF70664.1; --
 DR EMBL; AF248231; AAF70665.1; --
 DR EMBL; AF248232; AAF70666.1; --
 DR EMBL; AF248233; AAF70667.1; --
 DR EMBL; AF248234; AAF70668.1; --
 DR EMBL; AB007458; AAK75729.1; --
 DR EMBL; AB008518; AAL00298.1; --
 DR PIR; E98058; E98058.
 DR PIR; H95191; H95191.
 DR PDB; 1PSZ; 19-APR-00.
 DR TIGR; SP1650; --
 DR InterPro; IPR006128; Lipoprotein 4.
 DR InterPro; IPR000437; Prok_lipoprot_4.
 DR InterPro; IPR006127; SBP_bac_9.
 DR Pfam; PF01297; SBP_bac_9; 1.
 DR PRINTS; PR00690; ADHESNFAMILY.
 DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
 KW Transport; Manganese; Membrane; Lipoprotein; Signal;
 KW Complete proteome; 3D-structure; Palmitate.
 FT SIGNAL 1 19 PROBABLE.
 FT CHAIN 20 309 MANGANESE ABC TRANSPORTER SUBSTRATE-BINDING LIPOPROTEIN.
 FT LIPID 20 20 N-palmitoyl cysteine (Probable).
 FT LIPID 20 20 S-diacetylglycerol cysteine (Probable).
 FT VARIANT 8 8 L -> F (IN STRAIN NA-1508/92).
 FT VARIANT 9 9 V -> I (IN STRAIN NA-1064/97).
 FT VARIANT 14 14 A -> V (IN STRAINS NA-1064/97, NA-1383/97 AND NA-1508/92).
 FT VARIANT 16 16 I -> A (IN STRAINS NA-1064/97 AND NA-1383/97).
 FT VARIANT 16 16 I -> V (IN STRAIN NA-1508/92).
 FT VARIANT 27 28 TT -> AA (IN STRAINS NA-1064/97, NA-1383/97 AND NA-1508/92).
 FT VARIANT 30 30 G -> S (IN STRAIN NA-1064/97).
 FT VARIANT 62 62 I -> V (IN STRAIN NA-1383/97).
 FT VARIANT 81 81 E -> Q (IN STRAIN NA-1383/97).
 FT VARIANT 83 83 D -> E (IN STRAIN TIGR4).
 FT VARIANT 120 120 D -> E (IN STRAINS NA-1064/97, NA-1383/97 AND NA-1508/92).
 FT VARIANT 130 130 Q -> K (IN STRAINS NA-1064/97 AND NA-1508/92).
 FT VARIANT 148 148 I -> M (IN STRAIN NA-1383/97).
 FT VARIANT 164 164 N -> S (IN STRAIN NA-1383/97).
 FT VARIANT 187 189 SKD -> AKE (IN STRAIN NA-1383/97).
 FT VARIANT 193 193 K -> N (IN STRAINS NA-1064/97, NA-1383/97 AND NA-1508/92).
 FT VARIANT 207 207 A -> C (IN STRAIN NA-1383/97).
 FT VARIANT 234 234 E -> D (IN STRAIN NA-1383/97).
 FT VARIANT 248 248 V -> T (IN STRAIN NA-1383/97).
 FT VARIANT 285 285 Q -> E (IN STRAIN NA-1508/92).
 FT VARIANT 294 294 S -> N (IN STRAIN NA-1383/97).
 SQ SEQUENCE 309 AA; 34594 MW; BL25E7FE3DA6F67C CRC64;

Query Match 37.5%; Score 6; DB 1; Length 309;
 Best Local Similarity 100.0%; Pred. No. 7.4;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PNNKEF 13
 DB 163 PNNKEF 168

RESULT 4
 ID MTSA_STRTM STANDARD; PRT; 309 AA.
 AC Q9LSX0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Manganese ABC transporter substrate-binding lipoprotein precursor
 DE (Pneumococcal surface adhesin A).
 DE PSAA.
 GN Streptococcus mitis.
 OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=28037;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CIP 103335T / NCTC 12261;
 RX MEDLINE=21418906; PubMed=11527799;
 RA Jado I., Fenoll A., Casal J., Perez A.;
 RT "Identification of the psaA gene, coding for pneumococcal surface
 RT adhesin A, in viridans group streptococci other than Streptococcus
 RT pneumoniae";
 RL Clin. Diagn. Lab. Immunol. 8:895-898(2001).
 CC -!- FUNCTION: Part of an ATP-driven transport system for manganese.
 CC Also act as an adhesin which is involved on adherence to
 CC extracellular matrix. It is an important factor in pathogenesis
 CC and infection (By similarity).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 CC (By similarity).
 CC -!- SIMILARITY: Belongs to the bacterial solute-binding protein family
 CC 9. Lipoprotein receptor antigen (Lral) subfamily.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF248236; AAF64229.1; --
 DR HSP; E96116; ITOR.
 DR InterPro; IPR006128; Lipoprotein 4.
 DR InterPro; IPR000437; Prok_lipoprot_4.
 DR InterPro; IPR006127; SBP_bac_9.
 DR Pfam; PF01297; SBP_bac_9; 1.
 DR PRINTS; PR00690; ADHESNFAMILY.
 DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
 KW Transport; Manganese; Membrane; Lipoprotein; Signal; Palmitate.
 FT SIGNAL 1 19 PROBABLE.
 FT CHAIN 20 309 MANGANESE ABC TRANSPORTER SUBSTRATE-BINDING LIPOPROTEIN.
 FT LIPID 20 20 N-palmitoyl cysteine (Probable).
 FT LIPID 20 20 S-diacetylglycerol cysteine (Probable).
 SQ SEQUENCE 309 AA; 34595 MW; 112423C9F4873D25 CRC64;

Query Match 37.5%; Score 6; DB 1; Length 309;
 Best Local Similarity 100.0%; Pred. No. 7.4;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PNNKEF 13
 DB 163 PNNKEF 168

RESULT 5
 ID CAP_DICDI STANDARD; PRT; 464 AA.

P54654;
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last sequence update)
 15-MAR-2004 (Rel. 43, Last annotation update)
 ADENYL cyclase-associated protein (CAP).
 GN
 Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX3;
 RX MEDLINE=96228685; PubMed=8688557;
 RA Gottwald U., Brokamp R., Karakesisoglou I., Schleicher M.,
 RA Noegel A.A.;
 RT Identification of a cyclase-associated protein (CAP) homologue in
 RT Dictyostelium discoideum and characterization of its interaction with
 RT actin.
 RL Mol. Biol. Cell 7:261-272(1996).
 CC AND PIP2 INVOLVED IN MICROFILAMENT REORGANIZATION NEAR THE PLASMA
 CC MEMBRANE IN A PIP2-REGULATED MANNER.
 CC -1- SUBCELLULAR LOCATION: Cell membrane.
 CC -1- DOMAIN: THE C-TERMINUS IS RESPONSIBLE FOR SEQUESTERING G-ACTIN.
 CC THE N-TERMINUS IS REQUIRED FOR THE PIP2 MODULATION OF CAP
 CC FUNCTION.
 CC -1- SIMILARITY: Belongs to the CAP family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U43027; AB09713.1; -;
 CC DictyBase; DDB0001743; CAP.
 CC InterPro; IPR001837; CAP.
 CC InterPro; IPR006599; CAP.
 CC Pfam; PF01213; CAP; 1.
 CC SMART; SM00673; CAP; 2.
 CC PROSITE; PS01089; CAP; 1; 1.
 CC PROSITE; PS01089; CAP; 2; 1.
 CC Membrane; Actin-binding.
 FT DOMAIN 33 42 POLY-SER.
 FT DOMAIN 217 253 PRO-RICH.
 FT SEQUENCE 464 AA; 49641 MW; 7ABDD05D8A0148B4 CRC64;
 Query Match 37.5%; Score 6; DB 1; Length 464;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 KEFKGV 16
 Db 178 KEFKGV 183
 RESULT 6
 PUR2_ARCFU STANDARD; PRT; 470 AA.
 AC 029108;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phosphoribosylamine-glycine ligase (EC 6.3.4.13) (GARS) (Glycinamide
 DE ribonucleotide synthetase) (Phosphoribosylglycinamide synthetase).
 GN PURD OR AF1157.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., Hickey E.K., Peterson J.D.,
 RA Richdum K.A., Dodson R.J., Gwinn M., Kyrpides N.C.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Sutton G.G., Gill S.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Adams M.D., Loftus B.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Badger J.H., Glodek A., Zhou L.,
 RA Peterson S., Reich C.I., McNeil L.K., Weidman J.P., McDonald L., Utterback T.,
 RA Overbeek R., Gocayne J.D., Weidman J.P., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.M., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997)
 CC -1- CATALYTIC ACTIVITY: ATP + 5-phospho-D-riboseylglycinamide.
 CC + phosphate + N(1)-(5-phospho-D-riboseyl)glycinamide.
 CC -1- PATHWAY: De novo purine biosynthesis; second step.
 CC -1- SIMILARITY: Belongs to the GARS family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB001024; AB900089.1; -;
 CC PIR; D69394; D69394.
 CC HSP; P15640; IGSO.
 CC TIGR; AF1157; -;
 CC HAMAP; MF_00138; atypical; 1.
 CC InterPro; IPR000115; Gars.
 CC Pfam; PF01071; GARS; 1.
 CC Pfam; PF02842; GARS B; 1.
 CC Pfam; PF02843; GARS C; 1.
 CC Pfam; PF02844; GARS N; 1.
 CC TIGRFAMs; TIGR00877; purd; 1.
 CC PROSITE; PS00184; GARS; FALSE NEG.
 CC Purine biosynthesis; Ligase; Complete proteome.
 FT DOMAIN 230 260 INSERT.
 FT SEQUENCE 470 AA; 51750 MW; C3B90CED22DCD353 CRC64;
 Query Match 37.5%; Score 6; DB 1; Length 470;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 KEFKGV 16
 Db 299 KEFKGV 304
 RESULT 7
 VTDB_MOUSE STANDARD; PRT; 476 AA.
 ID VTDB_MOUSE
 AC P21614; Q91XG1; Q9CY31;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Vitamin D-binding protein precursor (DBP) (Group-specific component)
 DE (Gc-globulin) (VDB).
 GN GC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic liver;
 RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuenli P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Marzarelli J., Mombaerts P.,
RA Nordone P., Ring H., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RC [2]
RC TISSUE=Kidney;
RC MEDLINE=2388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Diatchenko L., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.J., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RN SEQUENCE OF 5-476 FROM N.A.
RX MEDLINE=90353947; PubMed=1696927;
RA Yang F., Bergeron J.M., Linehan L.A., Lalley P.A., Sakaguchi A.Y.,
RA Bowman B.H.;
RT "Mapping and conservation of the group-specific component gene in
RT mouse.";
RL Genomics 7:509-516(1990).
RN [4]
RN SEQUENCE OF 17-38.
RX MEDLINE=89211545; PubMed=32433374;
RA Borke J.L., Litwiler R.D., Bell M.P., Pass D.N., McKean D.J.,
RA Kumar R.;
RT "The isolation, characterization and amino terminal sequence of the
RT vitamin D-binding protein (group specific component) from mouse
RT plasma.";
RL Int. J. Biochem. 20:1343-1349(1988).
CC -1- FUNCTION: Multifunctional protein found in plasma, ascitic fluid,
CC cerebrospinal fluid, and urine and on the surface of many cell
CC types. In plasma, it carries the vitamin D sterols and prevents
CC polymerization of actin by binding its monomers. DBP associates
CC with membrane-bound immunoglobulin on the surface of B-lymphocytes
CC and with IgG fc receptor on the membranes of T-lymphocytes.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the ALB/AFB/VDB family.
CC -1- SIMILARITY: Contains 3 albumin domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK010965; BAB27297.1; -;
DR EMBL; BC010762; AAH10762.1; -;
DR EMBL; M55413; AAA37669.1; -;
DR F01; A35327; A35327.
DR MGI; MGI:95669; Gc.
DR InterPro; IPR000264; Serum albumin.
DR Pfam; PF00273; transport_prot; 2.
DR PRINTS; PR00802; SERUMALBUMIN.
DR ProDom; PD002486; Serum albumin; 1.
DR SMART; SM00103; ALBUMIN; 2.
DR PROSITE; PS00212; ALBUMIN; 1.
KW Glycoprotein; Vitamin D; Transport; Actin-binding; Repeat; Signal.
FT SIGNAL 1 16
FT CHAIN 17 476 VITAMIN D-BINDING PROTEIN.
FT DOMAIN 20 202 ALBUMIN 1.
FT DOMAIN 208 388 ALBUMIN 2.
FT DOMAIN 395 476 ALBUMIN 3.
FT DISULFID 29 75 BY SIMILARITY.
FT DISULFID 74 83 BY SIMILARITY.
FT DISULFID 96 112 BY SIMILARITY.
FT DISULFID 111 122 BY SIMILARITY.
FT DISULFID 145 190 BY SIMILARITY.
FT DISULFID 189 198 BY SIMILARITY.
FT DISULFID 220 266 BY SIMILARITY.
FT DISULFID 265 273 BY SIMILARITY.
FT DISULFID 286 300 BY SIMILARITY.
FT DISULFID 299 311 BY SIMILARITY.
FT DISULFID 335 376 BY SIMILARITY.
FT DISULFID 375 384 BY SIMILARITY.
FT DISULFID 407 453 BY SIMILARITY.
FT DISULFID 452 462 BY SIMILARITY.
FT CONFLICT 247 247 N -> K (IN REF. 1).
SQ SEQUENCE 476 AA; 53600 MW; 633B0CE183CD43FD CRC64;
Query Match 37.5%; Score 6; DB 1; Length 476;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 STEEQV 7
DB 55 STEEQV 60
RESULT 8
VTDB RAT STANDARD; PRT; 476 AA.
AC P04276;
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vitamin D-binding protein precursor (DBP) (Group-specific component)
DE (Gc-globulin) (VDB).
GN GC OR DBP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91177870; PubMed=2007578;
RA Ray K., Wang X., Zhao M., Cooke N.E.;
RT "The rat vitamin D binding protein (Gc-globulin) gene. Structural
RT analysis, functional and evolutionary correlations.";
RL J. Biol. Chem. 266:6221-6229(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90094352; PubMed=2480956;
RP SEQUENCE 90094352; PubMed=2480956;
RX

RA McLeod J.F., Cooke N.E.;
 RT "The vitamin D-binding protein, alpha-fetoprotein, albumin multigene
 RL family: detection of transcripts in multiple tissues.";
 RL J. Biol. Chem. 264:21760-21769(1989).
 RP [3]
 RN SEQUENCE FROM N.A.
 RA MEDLINE=86140127; PubMed=2419332;
 RX Cooke N.E.;
 RT "Rat vitamin D binding protein. Determination of the full-length
 RL primary structure from cloned cDNA.";
 RL J. Biol. Chem. 261:3441-3450(1986).
 RN [4]
 RP SEQUENCE OF 17-40.
 RX MEDLINE=86229807; PubMed=3713442;
 RA Litwiler R.D., Fass D.N., Kumar R.;
 RT "The amino acid sequence of the NH₂-terminal portion of rat and human
 RL vitamin D binding protein: evidence for a high degree of homology
 RL between rat and human vitamin D binding protein.";
 RL Life Sci. 38:2179-2184(1986).
 CC -1- FUNCTION: Multifunctional protein found in plasma, ascitic fluid,
 CC cerebrospinal fluid, and urine and on the surface of many cell
 CC types. In plasma, it carries the vitamin D sterols and prevents
 CC polymerization of actin by binding its monomers. DAP associates
 CC with membrane-bound immunoglobulin on the surface of B-lymphocytes
 CC and with IgG Fc receptor on the membranes of T-lymphocytes.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to the ALB/AFP/VDB family.
 CC -1- SIMILARITY: Contains 3 albumin domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M60205; AAA41081.1; -;
 DR EMBL; M60197; AAA41081.1; JOINED.
 DR EMBL; M60198; AAA41081.1; JOINED.
 DR EMBL; M60199; AAA41081.1; JOINED.
 DR EMBL; M60200; AAA41081.1; JOINED.
 DR EMBL; M60201; AAA41081.1; JOINED.
 DR EMBL; M60202; AAA41081.1; JOINED.
 DR EMBL; M60203; AAA41081.1; JOINED.
 DR EMBL; M60204; AAA41081.1; JOINED.
 DR EMBL; M12450; AAA41080.1; -;
 DR EMBL; J05148; AAA41082.1; -;
 DR PIR; A38726; VIKTD.
 DR InterPro; IPR000264; Serum_albumin.
 DR Pfam; PF00273; transport_prot; 2.
 DR PRINTS; PR00802; SERUMALBUMIN.
 DR ProDom; PD002486; Serum_albumin; 1.
 DR SMART; SM00103; ALBUMIN; 2.
 DR PROSITE; PS00212; ALBUMIN; 1.
 KW Glycoprotein; Vitamin D; Transport; Actin-binding; Repeat; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 476 VITAMIN D-BINDING PROTEIN.
 FT DOMAIN 20 202 ALBUMIN 1.
 FT DOMAIN 208 388 ALBUMIN 2.
 FT DOMAIN 395 476 ALBUMIN 3.
 FT DISULFID 29 75 BY SIMILARITY.
 FT DISULFID 74 83 BY SIMILARITY.
 FT DISULFID 96 112 BY SIMILARITY.
 FT DISULFID 111 122 BY SIMILARITY.
 FT DISULFID 145 190 BY SIMILARITY.
 FT DISULFID 189 198 BY SIMILARITY.
 FT DISULFID 220 266 BY SIMILARITY.
 FT DISULFID 265 273 BY SIMILARITY.
 FT DISULFID 286 300 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 335 376 BY SIMILARITY.
 FT DISULFID 375 384 BY SIMILARITY.

FT DISULFID 407 453 BY SIMILARITY.
 FT DISULFID 452 462 BY SIMILARITY.
 FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 132 132 E -> Q (IN REF. 2).
 FT CONFLICT 174 174 L -> P (IN REF. 3).
 FT CONFLICT 210 210 L -> S (IN REF. 2).
 SQ SEQUENCE 476 AA; 53544 MW; D3C729BC44E221E CRC64;
 Query Match 37.5%; Score 6; DB 1; Length 476;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 STEFQV 7
 Db 55 STEFQV 60
 RESULT 9
 YSP3 YEAST STANDARD; PRT; 478 AA.
 ID YSP3 YEAST
 AC P25036;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Subtilisin-like protease III precursor (EC 3.4.24.-).
 GN YSP3 OR YOR003W.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mason O.B., Wong P.A., Barr P.J.;
 RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97051599; PubMed=8896276;
 RA Sterky F., Holmberg A., Pettersson B., Uhlen M.;
 RT "The sequence of a 30 kb fragment on the left arm of chromosome XV
 RT from Saccharomyces cerevisiae reveals 15 open reading frames, five of
 RT which correspond to previously identified genes.";
 RL Yeast 12:1091-1095(1996).
 CC -1- SIMILARITY: Belongs to peptidase family S8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M77197; AAA35237.1; -;
 DR EMBL; U43491; AAC49482.1; -;
 DR EMBL; Z74911; CAA39191.1; -;
 DR PIR; S61986; S61986.
 DR HSSP; Q45670; IDBI.
 DR GeneOnline; 143591; -;
 DR MEROPS; S08.00A; -;
 DR SGD; S0005529; YSP3.
 DR InterPro; IPR000209; Peptidase_S8.
 DR InterPro; IPR009020; Protease_inhib.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Signal.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 478 SUBTILISIN-LIKE PROTEASE III.
 FT ACT_SITE 213 243 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 245 245 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT

FT ACT SITE 407 407 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CONFLICT 22 22 D -> E (IN REF. 1).
 FT CONFLICT 105 105 N -> D (IN REF. 1).
 FT CONFLICT 189 205 MISSING (IN REF. 1).
 FT CONFLICT 289 Y -> L (IN REF. 1).
 SQ SEQUENCE 478 AA; 52089 MW; 3EAB71360235C6FF CRC64;

Query Match 37.5%; Score 6; DB 1; Length 478;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 NKEFKG 15
 DB 302 NKEFKG 307

RESULT 10
 VN1 HUMAN
 ID VN1 HUMAN STANDARD; PRT; 513 AA.
 AC Q95497; Q9UFI6; Q9UFP4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Pantetheinase precursor (EC 3.5.1.-) (pantetheine hydrolase)
 DE (Vascular non-inflammatory molecule 1) (Vanin 1) (Tiff66).
 GN VN1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=99003335; PubMed=9790769;
 RA Galland F., Malerque F., Bazin H., Mattei M.-G., Aurrand-Lions M.,
 RA Theillet C., Naquet P.;
 RT "Two human genes related to murine vanin-1 are located on the long arm
 of human chromosome 6";
 RL Genomics 53:203-213 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Prehn S., Friedrichson T., Henske A., Boehm S., Hartmann E.,
 RA Kurzhallia T.V.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Phillips S.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP FUNCTION.
 RX MEDLINE=20036435; PubMed=10567687;
 RA Maras B., Barra D., Dupre S., Pitari G.;
 RT "Is pantetheinase the actual identity of mouse and human vanin-1
 proteins?";
 RL FEBS Lett. 461:149-152 (1999).

CC -1- FUNCTION: Amidohydrolase that hydrolyzes specifically one of the
 CC carboxamide linkages in D-pantetheine thus recycling pantothenic
 CC acid (vitamin B5) and releasing cysteamine.
 CC -1- CATALYTIC ACTIVITY: D-pantetheine + H(2)O = D-pantothenate +
 CC cysteamine.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -1- TISSUE SPECIFICITY: Expressed in spleen, thymus, peripheral blood
 CC lymphocytes, and small intestine.
 CC -1- SIMILARITY: Belongs to the CN hydrolase family. BTD/VNN subfamily.
 CC -1- SIMILARITY: Contains 1 CN hydrolase domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; U39664; AAF21453.1; -;
 CC EMBL; AJ132099; CAA10568.1; -;
 CC EMBL; AL032821; CAB40075.1; -;
 CC Genew; HGNC:12705; VNN1.
 CC MIM; 603570; -;
 CC GO; GO:0006928; P:cell motility; TAS.
 CC InterPro; IPR003010; Ntlse/CNhydase.
 CC Pfam; PF00795; CN_hydrolase; 1.
 CC PROSITE; PS50263; CN_HYDROLASE; 1.
 CC KW Hydrolase; Signal; Glycoprotein; GPI-anchor; Lipoprotein.
 CC SIGNAL 1 21
 CC CHAIN 22 491
 CC PROPEP 492 513
 CC DOMAIN 25 328
 CC LIPID 491 491
 CC CARBOHYD 38 38
 CC CARBOHYD 130 130
 CC CARBOHYD 200 200
 CC CARBOHYD 283 283
 CC CARBOHYD 315 315
 CC CARBOHYD 353 353
 CC CONFLICT 26 26
 CC CONFLICT 423 423
 CC SEQUENCE 513 AA; 57023 MW; 8PEF5230A358675A CRC64;

Query Match 37.5%; Score 6; DB 1; Length 513;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 NKEFKG 15
 DB 330 NKEFKG 335

RESULT 11
 RE3_VIBPA STANDARD; PRT; 529 AA.
 ID RE3_VIBPA
 AC Q87W18,
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Peptide chain release factor 3 (RF-3).
 GN PRFC OR VP2440.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 RT distinct from that of V. cholerae";
 RL Lancet 361:743-749 (2003).
 CC -1- FUNCTION: Increases the formation of ribosomal termination
 CC complexes and stimulates activities of RF-1 and RF-2. It binds
 CC guanine nucleotides and has strong preference for UGA stop codons.
 CC It may interact directly with the ribosome. The stimulation of RF-
 CC 1 and RF-2 is significantly reduced by GTP and GDP, but not by GMP
 CC (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to the GTP-binding elongation factor family.
 CC PRFC subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF005081; BAC60703.1; -;
 DR HAMAP; MF 00072; -; 1.
 DR InterPro; IPR000795; EF GTPbind.
 DR InterPro; IPR009022; EFG III V.
 DR InterPro; IPR004161; EFTU D2.
 DR InterPro; IPR009000; Translat_factor.
 DR Pfam; PF00009; GTP_EFTU; 1.
 DR Pfam; PF03144; GTP_EFTU_D2; 1.
 DR PRINTS; PR00315; ELONGATNFCT.
 DR PROSITE; PS00301; EFATOR GTP; 1.
 KW Protein biosynthesis; GTP-binding; Complete proteome.
 FT NP_BIND 20 27 GTP (BY SIMILARITY).
 FT NP_BIND 88 92 GTP (BY SIMILARITY).
 FT NP_BIND 142 145 GTP (BY SIMILARITY).
 SQ SEQUENCE 529 AA; 59245 MW; 3BC032BD9CA9821C CRC64;

Query Match 37.5%; Score 6; DB 1; Length 529;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
 |||||
 Db 177 KEFKGV 182

RESULT 12

RF3_VIBV ID RF3_VIBV STANDARD; PRT; 529 AA.
 AC QRD6;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Peptide chain release factor 3 (RF-3).
 GN PRFC OR VV11721.
 OS Vibrio vulnificus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=672;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CMCP6;
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
 RA Choy H.E.;
 RT "Complete genome sequence of Vibrio vulnificus CMCP6."
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Increases the formation of ribosomal termination
 CC complexes and stimulates activities of RF-1 and RF-2. It binds
 CC guanine nucleotides and has strong preference for UGA stop codons.
 CC It may interact directly with the ribosome. The stimulation of RF-
 CC 1 and RF-2 is significantly reduced by GTP and GDP, but not by GMP
 CC (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to the GTP-binding elongation factor family.
 CC PrfC subfamily.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AE016802; AAC10136.1; -;
 DR HAMAP; MF 00072; -; 1.
 DR InterPro; IPR000795; EF GTPbind.
 DR InterPro; IPR009022; EFG III V.
 DR InterPro; IPR004161; EFTU D2.
 DR InterPro; IPR004548; PrfC.

DR InterPro; IPR005225; Small GTP.
 DR InterPro; IPR009000; Translat_factor.
 DR Pfam; PF00009; GTP_EFTU; 1.
 DR Pfam; PF03144; GTP_EFTU_D2; 1.
 DR PRINTS; PR00315; ELONGATNFCT.
 DR TIGRFAMs; TIGR00503; prfc; 1.
 DR TIGRFAMs; TIGR00231; small GTP; 1.
 DR PROSITE; PS00301; EFATOR GTP; 1.
 KW Protein biosynthesis; GTP-binding; Complete proteome.
 FT NP_BIND 20 27 GTP (BY SIMILARITY).
 FT NP_BIND 88 92 GTP (BY SIMILARITY).
 FT NP_BIND 142 145 GTP (BY SIMILARITY).
 SQ SEQUENCE 529 AA; 59197 MW; 970C3B52CB6070C4 CRC64;

Query Match 37.5%; Score 6; DB 1; Length 529;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
 |||||
 Db 177 KEFKGV 182

RESULT 13

RF3_VIBV ID RF3_VIBV STANDARD; PRT; 529 AA.
 AC Q7M34;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Peptide chain release factor 3 (RF-3).
 GN PRFC OR VV2683.
 OS Vibrio vulnificus (strain VV016).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=196600;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Chen C.Y., Wu K.M., Chang Y.C., Chang C.H., Tsai H.C., Liao T.L.,
 RA Liu Y.M., Chen H.J., Shen A.B., Li J.C., Su T.L., Shao C.P., Lee C.T.,
 RA Hor L.I., Tsai S.F.;
 RT "Comparative genome analysis of Vibrio vulnificus, a marine
 pathogen."
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Increases the formation of ribosomal termination
 CC complexes and stimulates activities of RF-1 and RF-2. It binds
 CC guanine nucleotides and has strong preference for UGA stop codons.
 CC It may interact directly with the ribosome. The stimulation of RF-
 CC 1 and RF-2 is significantly reduced by GTP and GDP, but not by
 CC GMP (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to the GTP-binding elongation factor family.
 CC PrfC subfamily.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AP005340; BAC95447.1; -;
 DR HAMAP; MF 00072; -; 1.
 DR PROSITE; PS00301; EFATOR GTP; 1.
 KW Protein biosynthesis; GTP-binding; Complete proteome.
 FT NP_BIND 20 27 GTP (BY SIMILARITY).
 FT NP_BIND 88 92 GTP (BY SIMILARITY).
 FT NP_BIND 142 145 GTP (BY SIMILARITY).
 SQ SEQUENCE 529 AA; 59223 MW; C6EBA67BA7B070D2 CRC64;

Query Match 37.5%; Score 6; DB 1; Length 529;
 Best Local Similarity 100.0%; Pred. No. 12;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
| | | | |
Db 177 KEFKGV 182

Query Match 37.5%; Score 6; DB 1; Length 531;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
| | | | |
Db 179 KEFKGV 184

RESULT 14
RF3_VIBCH STANDARD; PRT; 531 AA.
AC Q9KUG4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peptide chain release factor 3 (RF-3).
GN PRFC OR VC0659.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bae S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Frazer C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae";
RL Nature 406:477-483 (2000).
CC -!- FUNCTION: Increases the formation of ribosomal termination
CC complexes and stimulates activities of RF-1 and RF-2. It binds
CC guanine nucleotides and has strong preference for UGA stop codons.
CC It may interact directly with the ribosome. The stimulation of RF-
CC 1 and RF-2 is significantly reduced by GTP and GDP, but not by GMP
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the GTP-binding elongation factor family.
CC PRFC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB004152; AAF93825.1; -.
CC F1R; E82295; E82295.
CC TIGR; VC0659; -.
CC HAMAP; MF 00072; -; 1.
CC InterPro; IPR000795; EF_GTPbind.
CC InterPro; IPR009022; EFG_III_V.
CC InterPro; IPR004161; EFTU_D2.
CC InterPro; IPR004548; PRFC.
CC InterPro; IPR005225; Small GTP.
CC InterPro; IPR009000; Translat_factor.
CC Pfam; PF00009; GTP_EFTU_1.
CC Pfam; PF03144; GTP_EFTU_D2; 1.
CC PRINTS; PR00315; ELONGANFCT.
CC TIGRFAMs; TIGR00503; PRFC; 1.
CC TIGRFAMs; TIGR00231; small GTP; 1.
CC PROSITE; PS00301; EFACTOR GTP; 1.
CC Protein biosynthesis; GTP-binding; Complete proteome.
KW NP_BIND 22 29 GTP (BY SIMILARITY).
FT NP_BIND 90 94 GTP (BY SIMILARITY).
FT NP_BIND 144 147 GTP (BY SIMILARITY).
FT SEQUENCE 531 AA; 59626 MW; 794035854ACE5568 CRC64;

Query Match 37.5%; Score 6; DB 1; Length 531;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
| | | | |
Db 179 KEFKGV 184

RESULT 15
YL66 YEAST STANDARD; PRT; 701 AA.
AC Q06149;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative transcriptional regulatory protein YLR266C.
GN YLR266C OR L8479.13.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albertmann K., Andre B., Ansorge W.,
RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miolaga T., Moestl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Rechmann S., Rieger C., Rinke M., Rose M.,
RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestazu L.A., Vandenbol M., Verhaaselt P.,
RA Vierdeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL Nature 387:87-90 (1997).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Contains 1 Zn(2)-Cys(6) fungal-type binuclear cluster
CC domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U17244; AAB67383.1; -.
CC F1R; S51403; S51403.
CC HSP; P12351; 1PYC.
CC GERMOnline; 142328; -.
CC SGD; S0004256; YLR266C.
CC GO; GO:0005634; C:nucleus; IPI.
CC GO; GO:0003677; F:DNA binding; IDA.
CC GO; GO:0045944; P:positive regulation of transcription from P...; IGI.
CC GO; GO:0006950; P:response to stress; IGI.
CC InterPro; IPR007219; Fungal trans.
CC InterPro; IPR001138; Fungi_Trn.
CC Pfam; PF04082; Fungal_trans; 1.
CC Pfam; PF00172; Zn_c1us; 1.
CC PRINTS; PR00054; FUNGALZNCYS.
CC SMART; SM00066; GAL4; 1.
CC PROSITE; PS00463; ZN2_CV6_FUNGAL_1; 1.
CC PROSITE; PS00048; ZN2_CV6_FUNGAL_2; 1.
CC Hypothetical protein; Transcription regulation; DNA-binding;
KW Nuclear protein; Zinc; Metal-binding.
FT DNA_BIND 31 59 ZN(2)-CYS(6), FUNGAL-TYPE.
FT DOMAIN 272 278 POLY-SER.

SQ SEQUENCE 701 AA; 81273 MW; 78C8A82BDFD65F0E CRC64;
 Query Match 37.5%; Score 6; DB 1; Length 701;
 Best Local Similarity 100.0%; Pred.No. 15;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 FEQVFN 9
 |||||
 Db 75 FEQVFN 80

Search completed: April 23, 2004, 14:52:14
 Job time : 9 secs

GenCore version 5.1.16
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:47:36 ; Search time 33.3333 seconds
(without alignments)
151.449 Million cell updates/sec

Title: US-09-528-682-3_COPY_64_79
Perfect score: 16
Sequence: 1 MSTFRQVFNKKEFGV 16

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rivir.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6	37.5	37	10 Q948W3	Q948W3 physcomitre
2	6	37.5	108	16 Q8RR34	Q8RR34 synechococc
3	6	37.5	125	11 Q99KU9	Q99KU9 mus musculu
4	6	37.5	141	2 Q8VWX3	Q8VWX3 streptococc
5	6	37.5	141	2 Q8VL74	Q8VL74 streptococc
6	6	37.5	141	2 Q8VLH4	Q8VLH4 streptococc
7	6	37.5	141	2 Q8VLR6	Q8VLR6 streptococc
8	6	37.5	141	2 Q8VLG2	Q8VLG2 streptococc
9	6	37.5	141	2 Q8VXX6	Q8VXX6 streptococc
10	6	37.5	141	2 Q8VL76	Q8VL76 streptococc
11	6	37.5	141	2 Q8VL85	Q8VL85 streptococc
12	6	37.5	141	2 Q8VL87	Q8VL87 streptococc
13	6	37.5	141	2 Q8VXX5	Q8VXX5 streptococc
14	6	37.5	141	2 Q8VXX8	Q8VXX8 streptococc
15	6	37.5	141	2 Q8VXX9	Q8VXX9 streptococc
16	6	37.5	141	2 Q8VXX4	Q8VXX4 streptococc

17	6	37.5	141	2	Q8VL40	Q8VL40 streptococc
18	6	37.5	141	2	Q8VXX7	Q8VXX7 streptococc
19	6	37.5	153	2	Q31302	Q31302 corynebacte
20	6	37.5	205	16	Q9CJW2	Q9CJW2 pasteurellia
21	6	37.5	208	16	Q8XJB0	Q8XJB0 clostridium
22	6	37.5	283	16	Q81DT1	Q81DT1 bacillus ce
23	6	37.5	295	16	Q97PB9	Q97PB9 streptococc
24	6	37.5	295	16	Q8DNS9	Q8DNS9 streptococc
25	6	37.5	296	16	Q8Y8V0	Q8Y8V0 listeria mo
26	6	37.5	309	2	Q8VQ82	Q8VQ82 streptococc
27	6	37.5	309	16	Q87G80	Q87G80 vibrio para
28	6	37.5	376	16	Q8R799	Q8R799 thermomanaer
29	6	37.5	413	16	Q883B6	Q883B6 pseudomonas
30	6	37.5	454	16	Q8A601	Q8A601 bacteroides
31	6	37.5	466	8	Q32351	Q32351 gelidium sp
32	6	37.5	490	5	Q7YIC9	Q7YIC9 cryptospori
33	6	37.5	491	11	Q7TS97	Q7TS97 mus musculu
34	6	37.5	499	16	Q9PQX1	Q9PQX1 ureaplasma
35	6	37.5	512	3	Q874I7	Q874I7 candida tro
36	6	37.5	512	3	Q874I6	Q874I6 candida tro
37	6	37.5	515	16	Q8RG77	Q8RG77 fusobacteri
38	6	37.5	540	16	Q8RA15	Q8RA15 thermomanaer
39	6	37.5	547	5	Q95RP4	Q95RP4 throsophila
40	6	37.5	565	16	O51684	O51684 borrelia bu
41	6	37.5	604	16	Q8EWC2	Q8EWC2 mycoplasma
42	6	37.5	663	16	O31529	O31529 bacillus su
43	6	37.5	727	2	Q53724	Q53724 staphylococ
44	6	37.5	785	16	Q7VN79	Q7VN79 haemophilus
45	6	37.5	836	16	Q7UNF7	Q7UNF7 rhodospirell

ALIGNMENTS

RESULT 1

Q948W3 PRELIMINARY; PRT; 37 AA.
 ID Q948W3
 AC Q948W3;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Gene, complete cds, clone:KC4, similar to pol polyprotein.
 OS Physcomitrella patens (Moss).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
 OC Bryopsida; Funariaceae; Funariales; Funariaceae; Physcomitrella.
 OX NCBI_TaxID=3218;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Akita M., Valkonen J.P.T.;
 RT "Physcomitrella patens genomic DNA, putative retrotransposon, KC4.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DDSI databases.
 DR EMBL; AB048267; BAB64355.1;
 DR GO; GO:0000785; C:chromatin; IEA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003682; F:chromatin binding; IEA.
 DR GO; GO:0006333; P:chromatin assembly/disassembly; IEA.
 DR InterPro; IPR000953; Chromo.
 DR PROSITE; PS50013; CHROMO_2; 1.
 KW Polyprotein.
 SQ SEQUENCE 37 AA; 4718 MW; B73D0D3A67AB68DE CRC64;

Query Match 37.5%; Score 6; DB 10; Length 37;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 NNKEFK 14
 |||||
 Db 3 NNKEFK 8

RESULT 2

Q8RR34 PRELIMINARY; PRT; 108 AA.
 ID Q8RR34


```

AC Q9RR34;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Circadian clock protein KaiB.
GN KAIB OR TLR0482.
OS Synecococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RP Uzumaki T., Hayashi F., Onai K., Ishiura M.;
RT "Circadian clock gene cluster kaiABC in Synecococcus elongatus."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RK MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matano A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1."
RL DNA Res. 9:123-130(2002).
DR EMBL; AB071375; BAB85984.1; -.
DR EMBL; AP005370; BAC08034.1; -.
KW Complete proteome.
SQ SEQUENCE 108 AA; 12025 MW; BC6B71FBE6254E3 CRC64;

Query Match 37.5%; Score 6; DB 16; Length 108;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
Db 34 KEFKGV 39

RESULT 3
Q99KU9 PRELIMINARY; PRT; 125 AA.
AC Q99KU9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN NSAP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004001; AA040001.1; -.
DR MGD; MGI:1891690; Nsap1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 125 AA; 14117 MW; 790529CB52A2BFEE CRC64;

Query Match 37.5%; Score 6; DB 11; Length 125;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 VPKNKE 12
Db 102 VPKNKE 107

RESULT 4
Q8VVX3

```

```

ID Q8VVX3 PRELIMINARY; PRT; 141 AA.
AC Q8VVX3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fructokinase (Fragment).
GN SCRK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson D.A., Briles D.E., Crain M.J., Hollingshead S.K.;
RT "The Evolution and Virulence of Serogroup 6 Pneumococci on a Global
RT Scale."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF436775; AAL31012.1; -.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR000600; ROK.
DR Pfam; PF00480; ROK; 1.
DR PROSITE; PS01125; ROK; 1.
KW Kinase.
FT NON_TER 1
FT NON_TER 141
SQ SEQUENCE 141 AA; 15198 MW; 8620770FD17A1D51 CRC64;

Query Match 37.5%; Score 6; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
Db 84 KEFKGV 89

RESULT 5
Q8VL74 PRELIMINARY; PRT; 141 AA.
AC Q8VL74;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fructokinase (Fragment).
GN SCRK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson D.A., Briles D.E., Crain M.J., Hollingshead S.K.;
RT "The Evolution and Virulence of Serogroup 6 Pneumococci on a Global
RT Scale."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF436746; AAL30983.1; -.
DR EMBL; AF436754; AAL30991.1; -.
DR EMBL; AF436773; AAL31010.1; -.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR000600; ROK.
DR Pfam; PF00480; ROK; 1.
DR PROSITE; PS01125; ROK; 1.
KW Kinase.
FT NON_TER 1
FT NON_TER 141
SQ SEQUENCE 141 AA; 15238 MW; BD5674FCFAEBBFBDB CRC64;

Query Match 37.5%; Score 6; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
Db 84 KEFKGV 89

```

```
Db      84 KEFGV 89

RESULT 6
ID Q8VLH4 PRELIMINARY; PRT; 141 AA.
AC Q8VLH4;
DT 01-WAR-2002 (TREMBlrel. 20, Created)
DT 01-WAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Fructokinase (Fragment).
GN SCRK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson D.A., Briles D.E., Crain M.J., Hollingshead S.K.;
RT "The Evolution and Virulence of Serogroup 6 Pneumococci on a Global
RT Scale.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF436756; AAL30993.1; -.
DR EMBL; AF436768; AAL31005.1; -.
DR EMBL; AF436774; AAL31011.1; -.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR000600; ROK.
DR Pfam; PF00480; ROK; 1.
DR PROSITE; PS01125; ROK; 1.
KW Kinase.
FT NON_TER 1 1
FT NON_TER 141 141
SQ SEQUENCE 141 AA; 15193 MW; 4618C013FAED1FDD CRC64;

Query Match 37.5%; Score 6; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFGV 16
Db 84 KEFGV 89

RESULT 7
ID Q8VL86 PRELIMINARY; PRT; 141 AA.
AC Q8VL86;
DT 01-WAR-2002 (TREMBlrel. 20, Created)
DT 01-WAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Fructokinase (Fragment).
GN SCRK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson D.A., Briles D.E., Crain M.J., Hollingshead S.K.;
RT "The Evolution and Virulence of Serogroup 6 Pneumococci on a Global
RT Scale.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF436743; AAL30980.1; -.
DR EMBL; AF436745; AAL30982.1; -.
DR EMBL; AF436749; AAL30986.1; -.
DR EMBL; AF436751; AAL30988.1; -.
DR EMBL; AF436752; AAL30989.1; -.
DR EMBL; AF436762; AAL30999.1; -.
DR EMBL; AF436770; AAL31007.1; -.
DR EMBL; AF436772; AAL31009.1; -.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR000600; ROK.
DR Pfam; PF00480; ROK; 1.

RESULT 8
ID Q8VLG2 PRELIMINARY; PRT; 141 AA.
AC Q8VLG2;
DT 01-WAR-2002 (TREMBlrel. 20, Created)
DT 01-WAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Fructokinase (Fragment).
GN SCRK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson D.A., Briles D.E., Crain M.J., Hollingshead S.K.;
RT "The Evolution and Virulence of Serogroup 6 Pneumococci on a Global
RT Scale.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF436748; AAL30985.1; -.
DR EMBL; AF436761; AAL30998.1; -.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR000600; ROK.
DR Pfam; PF00480; ROK; 1.
DR PROSITE; PS01125; ROK; 1.
KW Kinase.
FT NON_TER 1 1
FT NON_TER 141 141
SQ SEQUENCE 141 AA; 15223 MW; 585026F21AFAEAC1 CRC64;

Query Match 37.5%; Score 6; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFGV 16
Db 84 KEFGV 89

RESULT 9
ID Q8VX6 PRELIMINARY; PRT; 141 AA.
AC Q8VX6;
DT 01-WAR-2002 (TREMBlrel. 20, Created)
DT 01-WAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Fructokinase (Fragment).
GN SCRK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson D.A., Briles D.E., Crain M.J., Hollingshead S.K.;
RT "The Evolution and Virulence of Serogroup 6 Pneumococci on a Global
RT Scale.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
```

```

DR EMBL; AF436759; AAL30996.1; -.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR000600; ROK.
DR Pfam; PF00480; ROK; 1.
DR PROSITE; PS01125; ROK; 1.
DR KINASE.
FT NON TER 1 1
FT NON TER 141 141
SQ SEQUENCE 141 AA; 15253 MW; BD52DA56E584AFDB CRC64;

Query Match 37.5%; Score 6; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
DB 84 KEFKGV 89

RESULT 10
Q8VL76 PRELIMINARY; PRT; 141 AA.
AC Q8VL76;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fructokinase (Fragment).
GN SCRK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson D.A., Briles D.E., Crain M.J., Hollingshead S.K.;
RT "The Evolution and Virulence of Serogroup 6 Pneumococci on a Global
Scale.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF436741; AAL30978.1; -.
DR EMBL; AF436750; AAL30987.1; -.
DR EMBL; AF436765; AAL31002.1; -.
DR EMBL; AF436767; AAL31004.1; -.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR000600; ROK.
DR Pfam; PF00480; ROK; 1.
DR PROSITE; PS01125; ROK; 1.
DR KINASE.
FT NON TER 1 1
FT NON TER 141 141
SQ SEQUENCE 141 AA; 15225 MW; BD4DD546FAEBBFBDB CRC64;

Query Match 37.5%; Score 6; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
DB 84 KEFKGV 89

RESULT 11
Q8VL85 PRELIMINARY; PRT; 141 AA.
AC Q8VL85;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fructokinase (Fragment).
GN SCRK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson D.A., Briles D.E., Crain M.J., Hollingshead S.K.;
RT "The Evolution and Virulence of Serogroup 6 Pneumococci on a Global
Scale.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF436741; AAL30978.1; -.
DR EMBL; AF436750; AAL30987.1; -.
DR EMBL; AF436765; AAL31002.1; -.
DR EMBL; AF436767; AAL31004.1; -.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR000600; ROK.
DR Pfam; PF00480; ROK; 1.
DR PROSITE; PS01125; ROK; 1.
DR KINASE.
FT NON TER 1 1
FT NON TER 141 141
SQ SEQUENCE 141 AA; 15225 MW; BD4DD546FAEBBFBDB CRC64;

Query Match 37.5%; Score 6; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
DB 84 KEFKGV 89

RESULT 12
Q8VL87 PRELIMINARY; PRT; 141 AA.
AC Q8VL87;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fructokinase (Fragment).
GN SCRK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson D.A., Briles D.E., Crain M.J., Hollingshead S.K.;
RT "The Evolution and Virulence of Serogroup 6 Pneumococci on a Global
Scale.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF436760; AAL30997.1; -.
DR EMBL; AF436776; AAL31013.1; -.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR000600; ROK.
DR Pfam; PF00480; ROK; 1.
DR PROSITE; PS01125; ROK; 1.
DR KINASE.
FT NON TER 1 1
FT NON TER 141 141
SQ SEQUENCE 141 AA; 15234 MW; B6496BECE584BFBDB CRC64;

Query Match 37.5%; Score 6; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
DB 84 KEFKGV 89

RESULT 13
Q8VX5 PRELIMINARY; PRT; 141 AA.
ID Q8VX5
AC Q8VX5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA Robinson D.A., Briles D.E., Crain M.J., Hollingshead S.K.;
RT "The Evolution and Virulence of Serogroup 6 Pneumococci on a Global
Scale.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF436753; AAL30990.1; -.
DR EMBL; AF436755; AAL30992.1; -.
DR EMBL; AF436758; AAL30995.1; -.
DR EMBL; AF436763; AAL31000.1; -.
DR EMBL; AF436764; AAL31001.1; -.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR000600; ROK.
DR Pfam; PF00480; ROK; 1.
DR PROSITE; PS01125; ROK; 1.
DR KINASE.
FT NON TER 1 1
FT NON TER 141 141
SQ SEQUENCE 141 AA; 15206 MW; B65664FCFAEBAFDB CRC64;

Query Match 37.5%; Score 6; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
DB 84 KEFKGV 89

RESULT 12
Q8VL87 PRELIMINARY; PRT; 141 AA.
AC Q8VL87;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fructokinase (Fragment).
GN SCRK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson D.A., Briles D.E., Crain M.J., Hollingshead S.K.;
RT "The Evolution and Virulence of Serogroup 6 Pneumococci on a Global
Scale.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF436760; AAL30997.1; -.
DR EMBL; AF436776; AAL31013.1; -.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR000600; ROK.
DR Pfam; PF00480; ROK; 1.
DR PROSITE; PS01125; ROK; 1.
DR KINASE.
FT NON TER 1 1
FT NON TER 141 141
SQ SEQUENCE 141 AA; 15234 MW; B6496BECE584BFBDB CRC64;

Query Match 37.5%; Score 6; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
DB 84 KEFKGV 89

RESULT 13
Q8VX5 PRELIMINARY; PRT; 141 AA.
ID Q8VX5
AC Q8VX5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)

```

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fructokinase (Fragment).
GN SCRK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson D.A., Briles D.E., Crain M.J., Hollingshead S.K.;
RT "The Evolution and Virulence of Serogroup 6 Pneumococci on a Global
RT Scale."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF436766; AAL31003.1; -.
DR GO; GO:0016301; F.kinase activity; IEA.
DR InterPro; IPR000600; ROK.
DR Pfam; PF00480; ROK; 1.
DR PROSITE; PS01125; ROK; 1.
KW Kinase.
FT NON_TER 1 1
FT NON_TER 141 141
SQ SEQUENCE 141 AA; 15241 MW; BD4DC006FAFEFFDB CRC64;

Query Match 37.5%; Score 6; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
Db 84 KEFKGV 89

RESULT 14

Q8VVK8 PRELIMINARY; PRT; 141 AA.
ID Q8VVK8
AC Q8VVK8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fructokinase (Fragment).
GN SCRK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson D.A., Briles D.E., Crain M.J., Hollingshead S.K.;
RT "The Evolution and Virulence of Serogroup 6 Pneumococci on a Global
RT Scale."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF436747; AAL30984.1; -.
DR GO; GO:0016301; F.kinase activity; IEA.
DR InterPro; IPR000600; ROK.
DR Pfam; PF00480; ROK; 1.
DR PROSITE; PS01125; ROK; 1.
KW Kinase.
FT NON_TER 1 1
FT NON_TER 141 141
SQ SEQUENCE 141 AA; 15165 MW; ED57D54D3DE60900 CRC64;

Query Match 37.5%; Score 6; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
Db 84 KEFKGV 89

RESULT 15

Q8VVK9

ID Q8VVK9 PRELIMINARY; PRT; 141 AA.
AC Q8VVK9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fructokinase (Fragment).
GN SCRK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Robinson D.A., Briles D.E., Crain M.J., Hollingshead S.K.;
RT "The Evolution and Virulence of Serogroup 6 Pneumococci on a Global
RT Scale."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF436742; AAL30979.1; -.
DR GO; GO:0016301; F.kinase activity; IEA.
DR InterPro; IPR000600; ROK.
DR Pfam; PF00480; ROK; 1.
DR PROSITE; PS01125; ROK; 1.
KW Kinase.
FT NON_TER 1 1
FT NON_TER 141 141
SQ SEQUENCE 141 AA; 15227 MW; 1C18C013FAED12B0 CRC64;

Query Match 37.5%; Score 6; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
Db 84 KEFKGV 89

Search completed: April 23, 2004, 14:54:09
Job time : 35.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:45:50 ; Search time 48 Seconds
(without alignments)
94.183 Million cell updates/sec

Title: US-09-528-682-3_COPY_64_79
Perfect score: 16
Sequence: 1 MSTERQVNNKEFGV 16

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	6	37.5	77	4	AAG74544 Human col
2	6	37.5	124	4	AAM13825 Peptide #
3	6	37.5	124	4	ABB32770 Peptide #
4	6	37.5	124	4	AAM26232 Peptide #
5	6	37.5	124	4	ABB27600 Human pep
6	6	37.5	124	4	ABB18253 Protein #
7	6	37.5	124	4	AAM65958 Human bon
8	6	37.5	124	4	AAM53577 Human bra
9	6	37.5	124	4	ABG47622 Human liv
10	6	37.5	124	4	AAM01569 Peptide #
11	6	37.5	124	5	ABG35604 Human pep
12	6	37.5	191	4	ABG17164 Novel hum
13	6	37.5	245	4	AAM66543 Propionib
14	6	37.5	245	6	ABM63062 Propionib
15	6	37.5	270	6	ABM65318 Propionib
16	6	37.5	289	2	AAM55072 Streptoco
17	6	37.5	289	5	ABP54566 S. pneumo
18	6	37.5	289	7	ADC45101 S. pneumo
19	6	37.5	295	6	ABU02197 S. pneumo
20	6	37.5	296	5	ABM49395 Listeria
21	6	37.5	296	6	ABU24247 Protein e
22	6	37.5	309	2	AAM82496 S. pneumo
23	6	37.5	309	2	AAY30350 37 kDa pn
24	6	37.5	309	4	AAM01906 Streptoco
25	6	37.5	309	5	AAE19238 S. pneumo

ALIGNMENTS

RESULT 1
AAG74544
ID AAG74544 standard; protein; 77 AA.

XX AC AAG74544;
XX DT 03-SEP-2001 (first entry)
XX DE Human colon cancer antigen protein SEQ ID NO:5308.
XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX KW colorectal carcinoma; chromosome 20.
XX OS Homo sapiens.
XX PN WO200122920-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WO-US026524.
XX PR 29-SEP-1999; 99US-0157137P.
XX PR 03-NOV-1999; 99US-0163280P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX MPI; 2001-235357/24.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers.
XX Claim 11; Page 6963-6964; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where the
XX proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene therapy
XX and vaccine production. N and P may be used in the prevention, diagnosis
XX and treatment of diseases associated with inappropriate P expression. For
XX example, N and P may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of P by expressing inactive proteins or to
XX supplement the patients' own production of P. Additionally, N may be used
XX to produce the colon cancer-associated P, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the proteins. N and P
XX can be used in the prevention, diagnosis and treatment of colorectal

Abu02119 S. pneumo
Adc72221 S. pneumo
Aay81668 Streptoco
Aag72389 Human OR-
Aag71454 Human olf
Abu20374 Protein e
Aay90605 Candida t
Aay90605 Candida t
Aau12096 Candida t
Aau12097 Candida t
Adc44995 Yeast CYP
Adc44983 Yeast CYP
Adc44996 Yeast CYP
Adc44984 Yeast CYP
Adc45566 Yeast CYP
Adc45553 Yeast CYP
Adc45565 Yeast CYP
Adc45554 Yeast CYP
Ade52064 C. tropic
Ade52065 C. tropic

CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
 CC sequences used in the exemplification of the present invention. N.B.
 CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
 CC time of publication, meaning no sequences are present for SEQ ID NO:1027
 CC to 1052, 7921 and 7922

XX SQ Sequence 77 AA;

Query Match 37.5%; Score 6; DB 4; Length 77;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 VPNNKE 12
 |||||

DB 54 VPNNKE 59

RESULT 2

AA13825
 ID AAM13825 standard; protein; 124 AA.

XX AC AAM13825;

XX DT 12-OCT-2001 (first entry)

XX DE Peptide #259 encoded by probe for measuring cervical gene expression.

XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
 cervical cancer.

XX OS Homo sapiens.

XX FN WO200157278-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000670.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2001-488901/53.

XX DR Human genome-derived single exon nucleic acid probes useful for analyzing

XX PT gene expression in human cervical epithelial cells.

XX PS Claim 27; SEQ ID NO 18651; 487pp; English.

XX CC The present invention relates to human single exon nucleic acid probes
 (SENPs: see AAI10068-AA128459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 124 AA;

Query Match 37.5%; Score 6; DB 4; Length 124;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 VPNNKE 12
 |||||

DB 101 VPNNKE 106

RESULT 3

ABB32770
 ID ABB32770 standard; peptide; 124 AA.

XX AC ABB32770;

XX DT 04-FEB-2002 (first entry)

XX DE Peptide #276 encoded by human foetal liver single exon probe.

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.

XX OS Homo sapiens.

XX FN WO200157277-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000669.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2001-483447/52.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
 XX gene expression in human fetal liver.

XX PS Claim 27; SEQ ID NO 25405; 639pp + Sequence Listing; English.

XX CC The invention relates to a single exon nucleic acid probe for measuring
 human gene expression in a sample derived from human foetal liver. The
 CC single exon nucleic acid probes may be used for predicting, measuring and
 CC displaying gene expression in samples derived from human fetal liver. The
 CC present sequence is a peptide encoded by a single exon nucleic acid probe
 CC of the invention. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 124 AA;

Query Match 37.5%; Score 6; DB 4; Length 124;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 VPNNKE 12
 |||||

DB 101 VPNNKE 106

RESULT 4

AA26232
 ID AAM26232 standard; protein; 124 AA.

XX AC AAM26232;

XX DT 17-OCT-2001 (first entry)

XX XX

DE Peptide #269 encoded by probe for measuring placental gene expression.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200157272-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000663.
PF
XX
XX 04-FEB-2000; 2000US-0180312P.
PR
XX 26-MAY-2000; 2000US-0207456P.
PR
XX 30-JUN-2000; 2000US-00608408.
PR
XX 03-AUG-2000; 2000US-00632366.
PR
XX 21-SEP-2000; 2000US-0234687P.
PR
XX 27-SEP-2000; 2000US-0236359P.
PR
XX 04-OCT-2000; 2000US-00024263.
XX

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-488997/53.

Human genome-derived single exon nucleic acid probes useful for analyzing
gene expression in human placenta.

Claim 27; SEQ ID NO 26501; 654pp; English.

The present invention relates to single exon nucleic acid probes (SENP;
see AA13315-AA157546). The present sequence is a peptide encoded by one
such probe. The probes are useful for producing a microarray for
predicting, measuring and displaying gene expression in samples derived
from human placenta. The probes are useful for antenatal diagnosis of
human genetic disorders

Sequence 124 AA;

Query Match 37.5%; Score 6; DB 4; Length 124;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 VPNNKE 12
Db 101 VPNNKE 106

RESULT 5
ABB27600

ID ABB27600 standard; peptide; 124 AA.

AC ABB27600;

DT 01-FEB-2002 (first entry)

DE Human peptide #251 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast; disease;
cancer.

OS Homo sapiens.

XX WO200157271-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000662.

PF 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

PR

PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000US-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes, useful
for measuring gene expression in sample derived from human breast,
comprises number of single exon nucleic acid probes.
PT
XX Claim 27; SEQ ID NO 10568; 327pp + Sequence Listing; English.

The invention relates to a spatially-addressable set of single exon
nucleic acid probes for measuring gene expression in a sample derived
from human breast and BT 474 cells. The method involves contacting the
probes with a collection of detectably labelled nucleic acids derived
from mRNA of human breast, and then measuring the label bound to each
probe of the microarray. The probes are useful for verifying the
expression of regions of genomic DNA predicted to encode proteins. They
are useful for gene discovery, and for determining predisposition and/or
prognosing breast disease. Gene expression analysis is useful for
assessing the toxicity of chemical agents on cells. The microarray of
this invention presents a far greater diversity of probes for measuring
gene expression, with far less bias than expressed sequence tag
microarrays. The method is suitable for rapid production of functional
information from genomic sequence. The present sequence is a peptide
encoded by a single exon nucleic acid probe of the invention. Note: The
sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 124 AA;

Query Match 37.5%; Score 6; DB 4; Length 124;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 VPNNKE 12
Db 101 VPNNKE 106

RESULT 6
ABB18253

ID ABB18253 standard; protein; 124 AA.

AC ABB18253;

DT 23-JAN-2002 (first entry)

DE Protein #252 encoded by probe for measuring heart cell gene expression.
XX Human; gene expression; heart; microarray; vascular system;
cardiovascular disease; hypertension; cardiac arrhythmia;
congenital heart disease.

OS Homo sapiens.

XX WO200157274-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000665.

PF 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

```
PR 03-AUG-2000; 2000US-006323266.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.
XX
XX Claim 15; SEQ ID NO 20023; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart (see
XX ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX probe. The probes may be used for predicting, measuring and displaying
XX gene expression in samples derived from the human heart via microarrays.
XX By measuring gene expression, the probes are useful for predicting,
XX diagnosing, grading, staging, monitoring and prognosing diseases of the
XX human heart and vascular system e.g. cardiovascular disease,
XX hypertension, cardiac arrhythmias and congenital heart disease. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 124 AA;
XX
XX Query Match 37.5%; Score 6; DB 4; Length 124;
XX Best Local Similarity 100.0%; Pred. No. 59;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 7 VPNNKE 12
XX |||||
XX Db 101 VPNNKE 106
XX
XX RESULT 7
XX AAM65958
XX ID AAM65958 standard; protein; 124 AA.
XX
XX AC AAM65958;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 26264.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000668.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX
XX PR 26-MAY-2000; 2000US-0207456P.
XX
XX PR 30-JUN-2000; 2000US-00608408.
XX
XX PR 03-AUG-2000; 2000US-00632366.
XX
XX PR 21-SEP-2000; 2000US-0234687P.
XX
XX PR 27-SEP-2000; 2000US-0236359P.
XX
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
XX Example 4; SEQ ID NO 25682; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention
XX
```

```
DR WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX
XX PS Example 4; SEQ ID NO 26264; 659pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention
XX
XX SQ Sequence 124 AA;
XX
XX Query Match 37.5%; Score 6; DB 4; Length 124;
XX Best Local Similarity 100.0%; Pred. No. 59;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 7 VPNNKE 12
XX |||||
XX Db 101 VPNNKE 106
XX
XX RESULT 8
XX AAM53577
XX ID AAM53577 standard; protein; 124 AA.
XX
XX AC AAM53577;
XX
XX DT 05-NOV-2001 (first entry)
XX
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 25682.
XX
XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;
XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
XX OS Homo sapiens.
XX
XX PN WO200157275-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000667.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX
XX PR 26-MAY-2000; 2000US-0207456P.
XX
XX PR 30-JUN-2000; 2000US-00608408.
XX
XX PR 03-AUG-2000; 2000US-00632366.
XX
XX PR 21-SEP-2000; 2000US-0234687P.
XX
XX PR 27-SEP-2000; 2000US-0236359P.
XX
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
XX Example 4; SEQ ID NO 25682; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention
XX
```


Qy

AC ABG35604;
XX 19-AUG-2002 (first entry)
XX Human peptide encoded by genome-derived single exon probe SEQ ID 25269.
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX Homo sapiens.
XX WO200186003-A2.
XX 15-NOV-2001.
XX 30-JAN-2001; 2001WO-US000665.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples.
XX Claim 27; SEQ ID NO 25269; 634pp; English.
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of probes
CC; the novel set of probes which hybridise at high stringency to a nucleic
CC acid expressed in the human lung; measuring gene expression in a sample
CC derived from human lung, comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the
CC array; identifying exons in a eukaryotic genome, comprising (a)
CC algorithmically predicting at least one exon from genomic sequences of
CC the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene expression
CC analysis, and for identifying exons in a gene, particularly using human
CC lung derived mRNA and for the study of lung diseases such as asthma, lung
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-

CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
CC present sequence is a peptide/protein encoded by a single exon probe of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ffp.wipo.int/pub/published_pct_sequences
XX Sequence 124 AA;
SQ Sequence 124 AA;
Query Match 37.5%; Score 6; DB 5; Length 124;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 VPNNKE 12
Db 101 VPNNKE 106
RESULT 12
ABG17164
ID ABG17164 standard; protein; 191 AA.
XX AC ABG17164;
XX 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #17155.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS81351.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 47523; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG0377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 191 AA;

Query Match 37.5%; Score 6; DB 4; Length 191;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 NNKEK 14
 |||||
 DB 180 NNKEK 185

RESULT 13
 AAU6543
 ID AAU6543 standard; protein; 245 AA.
 XX AC
 XX AAU6543;
 XX 27-FEB-2002 (first entry)
 XX DE Propionibacterium acnes immunogenic protein #27439.
 XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX OS Propionibacterium acnes.
 XX FN WO200181581-A2.
 XX PD 01-NOV-2001.
 XX PF 20-APR-2001; 2001WO-US012865.
 XX PR 21-APR-2000; 2000US-0193047P.
 PR 02-JUN-2000; 2000US-0208841P.
 PR 07-JUL-2000; 2000US-0216747P.
 XX PA (CORI-) CORIXA CORP.
 XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 XX WPI: 2001-616774/71.
 DR N-PSDB; AAS59740.

XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.

XX Example 1; SEQ ID NO 27738; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and

CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 245 AA;

Query Match 37.5%; Score 6; DB 4; Length 245;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 STFEQV 7
 |||||
 DB 156 STFEQV 161

RESULT 14
 ABM63062
 ID ABM63062 standard; protein; 245 AA.
 XX AC
 XX ABM63062;
 XX 20-OCT-2003 (first entry)
 XX DE Propionibacterium acnes predicted ORF-encoded polypeptide #27738.
 XX KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;
 KW immunostimulant; immune response; vaccine.
 XX OS Propionibacterium acnes.
 XX FN WO2003033515-A1.
 XX PD 24-APR-2003.
 XX PF 11-OCT-2002; 2002WO-US032727.
 XX PR 15-OCT-2001; 2001US-00978825.
 XX PA (CORI-) CORIXA CORP.
 XX PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
 PI Barth B, Vallieve-Douglas J;
 XX WPI: 2003-381789/36.
 DR N-PSDB; ACF64669.

XX PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
 PT or for stimulating an immune response specific for a P. acnes protein.

XX Example 1; SEQ ID NO 27738; 1481pp; English.

XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
 CC encoding a Propionibacterium acnes protein. The invention also relates to
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
 CC immunogenic fragments of P. acnes polypeptides. The invention
 CC additionally encompasses expression vectors and host cells comprising a
 CC polynucleotide of the invention; antibodies against polypeptides of the
 CC invention; fusion proteins comprising a polypeptide of the invention; a
 CC method for stimulating an immune response specific for a P. acnes
 CC polypeptide and an isolated T cell population comprising T cells prepared
 CC via this method; and a vaccine composition (comprising P. acnes polypeptides,
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or
 CC antigen-presenting cells that express the polypeptide); a method and kit
 CC for detecting or determining the presence or absence of P. acnes in a
 CC patient; and a method for inhibiting the development of P. acnes in a
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations or antigen-presenting cells that express the
 CC polypeptides are useful for diagnosing, preventing or treating acne

CC vulgaris, or for stimulating an immune response specific for a P. acnes
 CC protein. The polynucleotides can also be used as probes or primers for
 CC nucleic acid hybridization. The vaccine composition is useful for the
 CC stimulation of an immune response against P. acnes, or for treating acne,
 CC and the kit is useful for performing a diagnostic assay. The present
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open
 CC reading frame) contained within the P. acnes polynucleotides of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 245 AA;

Query Match 37.5%; Score 6; DB 6; Length 245;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 STFEQV 7
 |||||
 Db 156 STFEQV 161

RESULT 15
 ABM65318
 ID ABM65318 standard; protein; 270 AA.

XX AC

XX ABM65318;

XX 20-OCT-2003 (first entry)

XX Propionibacterium acnes immunogenic polypeptide #29994.

XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;

XX immunostimulant; immune response; vaccine; immunogenic.

XX Propionibacterium acnes.

XX WO2003033515-A1.

XX 24-APR-2003.

XX 11-OCT-2002; 2002WO-US032727.

XX 15-OCT-2001; 2001US-00978825.

XX (CORI-) CORIXA CORP.

XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;

XX Zhang Y, Wang S, Jen S, Lodes MU, Benson DR, Jones R, Carter D;

XX Barth B, Valliave-Douglas J;

XX WPI; 2003-381789/36.

XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
 PT or for stimulating an immune response specific for a P. acnes protein.

XX Claim 7; SEQ ID NO 29994; 1481bp; English.

XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
 CC encoding a Propionibacterium acnes protein. The invention also relates to
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
 CC immunogenic fragments of P. acnes polypeptides. The invention
 CC additionally encompasses expression vectors and host cells comprising a
 CC polynucleotide of the invention; antibodies against polypeptides of the
 CC invention; fusion proteins comprising a polypeptide of the invention; a
 CC method for stimulating an immune response specific for a P. acnes
 CC polypeptide and an isolated T cell population comprising T cells prepared
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or
 CC antigen-presenting cells that express the polypeptide); a method and kit
 CC for detecting or determining the presence or absence of P. acnes in a
 CC patient; and a method for inhibiting the development of P. acnes in a

CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations or antigen-presenting cells that express the
 CC polypeptides are useful for diagnosing, preventing or treating acne
 CC vulgaris, or for stimulating an immune response specific for a P. acnes
 CC protein. The polynucleotides can also be used as probes or primers for
 CC nucleic acid hybridization. The vaccine composition is useful for the
 CC stimulation of an immune response against P. acnes, or for treating acne,
 CC and the kit is useful for performing a diagnostic assay. The present
 CC sequence represents a specifically claimed P. acnes polypeptide which is
 CC thought to contain an immunogenic region. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX

SQ Sequence 270 AA;

Query Match 37.5%; Score 6; DB 6; Length 270;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 STFEQV 7
 |||||
 Db 181 STFEQV 186

Search completed: April 23, 2004, 14:51:38
 Job time : 50 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:49:06 ; Search time 14 Seconds
(without alignments)
59.001 Million cell updates/sec

Title: US-09-528-682-3_COPY_64_79
Perfect score: 16
Sequence: 1 MSTFRQVFNKEFGV 16

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	50.0	241	US-08-823-120-1	Sequence 1, Appli
2	6	37.5	289	US-08-961-083-20	Sequence 20, Appl
3	6	37.5	289	US-09-536-784-20	Sequence 20, Appl
4	6	37.5	309	US-08-715-131-2	Sequence 2, Appli
5	6	37.5	309	US-09-221-753-2	Sequence 2, Appli
6	6	37.5	512	US-09-302-620B-102	Sequence 102, App
7	6	37.5	512	US-09-302-620B-103	Sequence 103, App
8	6	37.5	590	US-09-443-067-22	Sequence 22, Appl
9	5	31.2	27	US-08-615-942A-10	Sequence 10, Appl
10	5	31.2	27	US-09-237-325-10	Sequence 10, Appl
11	5	31.2	55	US-09-621-976-5400	Sequence 5400, Ap
12	5	31.2	101	US-09-025-596-50	Sequence 50, Appl
13	5	31.2	101	US-09-073-661-50	Sequence 50, Appl
14	5	31.2	101	US-10-100-785-50	Sequence 50, Appl
15	5	31.2	115	US-09-325-932A-51	Sequence 51, Appl
16	5	31.2	115	US-09-325-932A-52	Sequence 52, Appl
17	5	31.2	115	US-09-325-932A-53	Sequence 53, Appl
18	5	31.2	115	US-09-325-932A-54	Sequence 54, Appl
19	5	31.2	123	US-09-732-210-384	Sequence 384, App
20	5	31.2	130	US-09-107-532A-6322	Sequence 6322, Ap
21	5	31.2	148	US-09-370-838-215	Sequence 215, App
22	5	31.2	158	US-09-071-035-348	Sequence 348, App
23	5	31.2	161	US-09-634-238-255	Sequence 255, App
24	5	31.2	165	US-09-732-210-566	Sequence 566, App
25	5	31.2	171	US-09-543-681A-5644	Sequence 5644, Ap
26	5	31.2	195	US-09-071-035-346	Sequence 346, App
27	5	31.2	195	US-09-543-681A-7583	Sequence 7583, Ap

28 5 31.2 213 4 US-09-230-665-8 Sequence 8, Appli
29 5 31.2 229 4 US-09-546-043-8 Sequence 8, Appli
30 5 31.2 232 4 US-09-540-236-3762 Sequence 3762, Ap
31 5 31.2 240 4 US-09-107-532A-5420 Sequence 5420, Ap
32 5 31.2 250 1 US-08-248-466B-7 Sequence 7, Appli
33 5 31.2 257 3 US-09-183-217-2 Sequence 2, Appli
34 5 31.2 258 3 US-08-737-226-6 Sequence 6, Appli
35 5 31.2 260 1 US-08-015-985-10 Sequence 10, Appli
36 5 31.2 260 4 US-09-540-236-2236 Sequence 2236, Ap
37 5 31.2 260 4 US-09-280-597-10 Sequence 10, Appli
38 5 31.2 262 4 US-09-134-001C-5418 Sequence 5418, Ap
39 5 31.2 273 4 US-09-848-294-11 Sequence 11, Appl
40 5 31.2 303 4 US-09-495-406-17 Sequence 17, Appl
41 5 31.2 313 4 US-09-087-031E-4 Sequence 4, Appli
42 5 31.2 313 4 US-09-546-043-3 Sequence 3, Appli
43 5 31.2 313 4 US-09-514-885-1 Sequence 1, Appli
44 5 31.2 314 4 US-08-937-067-7 Sequence 7, Appli
45 5 31.2 314 4 US-09-087-031E-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-823-120-1
; Sequence 1, Application US/08823120
; Patent No. 6149919
; GENERAL INFORMATION:
; APPLICANT: Domenighini, Mario
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; TITLE OF INVENTION: Immunogenic Detoxified Mutants of
; TITLE OF INVENTION: Cholera toxin and of the toxin lct, Their Preparation and
; NUMBER OF INVENTION: Their Use for the Preparation of Vaccines
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,120
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,003
; FILING DATE: 11-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0315.001
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 241 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-823-120-1

Query Match 50.0%; Score 8; DB 3; Length 241;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      7 VPNNKEFK 14
Db      186 VPNNKEFK 193

RESULT 2
US-08-961-083-20
; Sequence 20, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; MISMATCHES: 0; Indels 0; Gaps 0;

Query Match      37.5%; Score 6; DB 3; Length 289;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 PNNKEF 13
Db      143 PNNKEF 148

RESULT 3
US-09-536-784-20
; Sequence 20, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

Query Match      37.5%; Score 6; DB 3; Length 289;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 PNNKEF 13
Db      143 PNNKEF 148

RESULT 4
US-08-715-131-2
; Sequence 2, Application US/08715131
; Patent No. 5854416
; GENERAL INFORMATION:
; APPLICANT: Sampson, Jacquelyn S.
; APPLICANT: Russell, Harold
; APPLICANT: Tharpe, Jean A.
; APPLICANT: Ades, Edwin W.
; APPLICANT: Carlone, George M.
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE 37-kDa SURFACE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303-1811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/715,131
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 14114.0200
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <UNKNOWN>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-536-784-20

Query Match      37.5%; Score 6; DB 4; Length 289;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 PNNKEF 13
Db      143 PNNKEF 148

RESULT 4
US-08-715-131-2
; Sequence 2, Application US/08715131
; Patent No. 5854416
; GENERAL INFORMATION:
; APPLICANT: Sampson, Jacquelyn S.
; APPLICANT: Russell, Harold
; APPLICANT: Tharpe, Jean A.
; APPLICANT: Ades, Edwin W.
; APPLICANT: Carlone, George M.
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE 37-kDa SURFACE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303-1811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/715,131
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 14114.0200
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770

```

```

; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 309 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   US-08-715-131-2
;
; Query Match          37.5%; Score 6; DB 2; Length 309;
; Best Local Similarity 100.0%; Pred. No. 30;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY      8 PNNKEF 13
; DB      163 PNNKEF 168
;
; RESULT 5
; US-09-221-753-2
;   Sequence 2, Application US/09221753
;   Patent No. 6217884
;   GENERAL INFORMATION:
;   APPLICANT: SAMPSON, JACQUELYN S.
;   APPLICANT: RUSSELL, HAROLD
;   APPLICANT: THARPE, JEAN A.
;   APPLICANT: ADES, EDWIN W.
;   APPLICANT: CARLONE, GEORGE M.
;   TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE 37 kDa SURFACE
;   FILE REFERENCE: 64778 US
;   CURRENT FILING DATE: 1998-12-28
;   EARLIER APPLICATION NUMBER: US 07/791,377
;   EARLIER FILING DATE: 1991-09-17
;   EARLIER APPLICATION NUMBER: US 07/816,286
;   EARLIER FILING DATE: 1992-01-03
;   EARLIER APPLICATION NUMBER: US 08/222,179
;   EARLIER FILING DATE: 1994-04-04
;   EARLIER APPLICATION NUMBER: US 08/715,131
;   EARLIER FILING DATE: 1996-09-17
;   NUMBER OF SEQ ID NOS: 4
;   SOFTWARE: FastSeq for Windows Version 3.0
;   SEQ ID NO 2
;   LENGTH: 309
;   TYPE: PRT
;   ORGANISM: STREPTOCOCCUS PNEUMONIAE
;   US-09-221-753-2
;
; Query Match          37.5%; Score 6; DB 3; Length 309;
; Best Local Similarity 100.0%; Pred. No. 30;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY      8 PNNKEF 13
; DB      163 PNNKEF 168
;
; RESULT 6
; US-09-302-620B-102
;   Sequence 102, Application US/09302620B
;   Patent No. 6331420
;   GENERAL INFORMATION:
;   APPLICANT: Wilson, C. Ron
;   APPLICANT: Craft, David L.
;   APPLICANT: Eirich, Dudley
;   APPLICANT: Eshoo, Mark
;   APPLICANT: Madduri, Krishna M.
;   APPLICANT: Cornett, Cathy A.
;   APPLICANT: Brenner, Alfred A.
;   APPLICANT: Tang, Maria
;   APPLICANT: Loper, John C.
;   APPLICANT: Gleeson, Martin
;
; Query Match          37.5%; Score 6; DB 4; Length 512;
; Best Local Similarity 100.0%; Pred. No. 46;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY      9 NNKEFK 14
; DB      246 NNKEFK 251
;
; RESULT 7
; US-09-302-620B-103
;   Sequence 103, Application US/09302620B
;   Patent No. 6331420
;   GENERAL INFORMATION:
;   APPLICANT: Wilson, C. Ron
;   APPLICANT: Craft, David L.
;   APPLICANT: Eirich, Dudley
;   APPLICANT: Eshoo, Mark
;   APPLICANT: Madduri, Krishna M.
;   APPLICANT: Cornett, Cathy A.
;   APPLICANT: Brenner, Alfred A.
;   APPLICANT: Tang, Maria
;   APPLICANT: Loper, John C.
;   APPLICANT: Gleeson, Martin
;   TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450
;   TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA
;   TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS
;   FILE REFERENCE: 1010-16.seq
;   CURRENT FILING DATE: 1999-04-30
;   CURRENT APPLICATION NUMBER: US/09/302,620B
;   NUMBER OF SEQ ID NOS: 109
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 103
;   LENGTH: 512
;   TYPE: PRT
;   ORGANISM: Candida tropicalis
;   US-09-302-620B-103
;
; Query Match          37.5%; Score 6; DB 4; Length 512;
; Best Local Similarity 100.0%; Pred. No. 46;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY      9 NNKEFK 14
; DB      246 NNKEFK 251
;
; RESULT 8
; US-09-443-067-22
;   Sequence 22, Application US/09443067
;   Patent No. 6627794
;   GENERAL INFORMATION:
;   APPLICANT: COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH
;   APPLICANT: ORGANISATION
;   TITLE OF INVENTION: Polyphenol oxidase genes from banana, lettuce, tobacco and
;   TITLE OF INVENTION: pineapple
;   FILE REFERENCE:

```

```
; CURRENT APPLICATION NUMBER: US/09/443,067
; CURRENT FILING DATE: 1999-11-18
; EARLIER APPLICATION NUMBER: US 08/976, 222
; EARLIER FILING DATE: 1997-11-21
; EARLIER APPLICATION NUMBER: PCT/AU98/00362
; EARLIER FILING DATE: 1998-05-19
; EARLIER APPLICATION NUMBER: AU PP3898
; EARLIER FILING DATE: 1995-05-23
; EARLIER APPLICATION NUMBER: AU PP6849
; EARLIER FILING DATE: 1997-05-19
; EARLIER APPLICATION NUMBER: AU PP5600
; EARLIER FILING DATE: 1995-09-26
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 590
; TYPE: PRT
; ORGANISM: banana
US-09-443-067-22

Query Match          37.5%; Score 6; DB 4; Length 590;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 TFEQVP 8
Db      408 TFEQVP 413

RESULT 9
US-06-615-942A-10
; Sequence 10, Application US/08615942A
; Patent No. 5863532
; GENERAL INFORMATION:
; APPLICANT: JOLINDA A. TRAUGH
; APPLICANT: REGINA D. ROONEY
; APPLICANT: ROLF JAKOBI
; APPLICANT: POLYGENA T. TUAZON
; APPLICANT: CHARNG-JUI CHEN
; APPLICANT: WILLIAM E. MEER
; APPLICANT: EDWARD J. CARROLL, JR.
; APPLICANT: CURTIS A. MONNIG
; TITLE OF INVENTION: Compositions and Methods Comprising
; TITLE OF INVENTION: Cytostatic Protein Kinase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Robbins, Berliner & Carson
; STREET: 201 N. Figueroa St., Suite 500
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90012
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/615,942A
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NONE
; FILING DATE: NONE
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Wong, Wean K.
; REGISTRATION NUMBER: 33,561
; REFERENCE/DOCKET NUMBER: 1279-203XX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213/977-1001
; TELEFAX: 213/977-1003
; INFORMATION FOR SEQ ID NO: 10:

; CURRENT APPLICATION NUMBER: US/09/443,067
; CURRENT FILING DATE: 1999-11-18
; EARLIER APPLICATION NUMBER: US 08/976, 222
; EARLIER FILING DATE: 1997-11-21
; EARLIER APPLICATION NUMBER: PCT/AU98/00362
; EARLIER FILING DATE: 1998-05-19
; EARLIER APPLICATION NUMBER: AU PP3898
; EARLIER FILING DATE: 1995-05-23
; EARLIER APPLICATION NUMBER: AU PP6849
; EARLIER FILING DATE: 1997-05-19
; EARLIER APPLICATION NUMBER: AU PP5600
; EARLIER FILING DATE: 1995-09-26
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 590
; TYPE: PRT
; ORGANISM: banana
US-09-443-067-22

Query Match          31.2%; Score 5; DB 2; Length 27;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      11 KEFKG 15
Db      16 KEFKG 20

RESULT 10
US-09-237-325-10
; Sequence 10, Application US/09237325
; Patent No. 6599726
; GENERAL INFORMATION:
; APPLICANT: Traugh, Jolinda A.
; APPLICANT: Rooney, Regina D.
; APPLICANT: Jakobi, Rolf
; APPLICANT: Tuazon, Polygena T.
; APPLICANT: Chen, Charnng-Jui
; APPLICANT: Meek, William E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS COMPRISING CYTOSTATIC PROTEIN
; FILE REFERENCE: 1279-203D1/986617
; CURRENT APPLICATION NUMBER: US/09/237,325
; CURRENT FILING DATE: 1999-01-25
; EARLIER APPLICATION NUMBER: 08/615,942
; EARLIER FILING DATE: 1996-03-14
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (3)
; OTHER INFORMATION: Author is unsure of the exact amino acid at this
; OTHER INFORMATION: position.
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (12)
; OTHER INFORMATION: Author is unsure of the exact amino acid at this
; OTHER INFORMATION: position.
US-09-237-325-10

Query Match          31.2%; Score 5; DB 4; Length 27;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      11 KEFKG 15
Db      16 KEFKG 20

RESULT 11
US-09-621-976-5400
; Sequence 5400, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
```



```

; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 5400
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -48...-1
US-09-621-976-5400

```

```

Query Match      31.2%; Score 5; DB 4; Length 55;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 3 TPEQV 7
Db 12 TPEQV 16

```

```

RESULT 12
US-09-025-596-50
; Sequence 50, Application US/09025596
; Patent No. 6340463
; GENERAL INFORMATION:
; APPLICANT: Mitchell, William M.
; APPLICANT: Stratton, Charles W.
; TITLE OF INVENTION: IDENTIFICATION OF ANTIGENIC PEPTIDE
; FILE REFERENCE: VDB98-01
; CURRENT APPLICATION NUMBER: US/09/025,596
; CURRENT FILING DATE: 1998-02-18
; EARLIER APPLICATION NUMBER: 08/911,593
; EARLIER FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 60/023,921
; EARLIER FILING DATE: 1996-08-14
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 50
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-025-596-50

```

```

Query Match      31.2%; Score 5; DB 4; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 10 NKEFK 14
Db 79 NKEFK 83

```

```

RESULT 13
US-09-073-661-50
; Sequence 50, Application US/09073661
; Patent No. 6579854
; GENERAL INFORMATION:
; APPLICANT: Mitchell, William M.
; APPLICANT: Stratton, Charles W.
; TITLE OF INVENTION: DIAGNOSIS AND MANAGEMENT OF
; TITLE OF INVENTION: INFECTION CAUSED BY CHLAMYDIA
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,661
; FILING DATE: 06-MAY-1998
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/025,521
; FILING DATE: 18-FEB-1998
; APPLICATION NUMBER: US 08/911,593
; FILING DATE: 14-AUG-1997
; APPLICATION NUMBER: US 60/023,921
; FILING DATE: 14-AUG-1996
; APPLICATION NUMBER: US 09/025,176
; FILING DATE: 18-FEB-1998
; APPLICATION NUMBER: US 09/025,174
; FILING DATE: 18-FEB-1998
; APPLICATION NUMBER: US 60/045,739
; FILING DATE: 6-MAY-1997
; APPLICATION NUMBER: US 60/045,779
; FILING DATE: 6-MAY-1997
; APPLICATION NUMBER: US 60/045,780
; FILING DATE: 6-MAY-1997
; APPLICATION NUMBER: US 60/045,784
; FILING DATE: 6-MAY-1997
; APPLICATION NUMBER: US 60/045,787
; FILING DATE: 6-MAY-1997
; APPLICATION NUMBER: US 60/045,689
; FILING DATE: 6-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Alice O
; REGISTRATION NUMBER: 33,542
; REFERENCE/DOCKET NUMBER: VDB97-09PM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 781-861-6240
; TELEFAX: 781-861-9540
; TELEX:
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 101 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-073-661-50

```

```

Query Match      31.2%; Score 5; DB 4; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 10 NKEFK 14
Db 79 NKEFK 83

```

```

RESULT 14
US-10-100-785-50
; Sequence 50, Application US/10100785
; Patent No. 6664239
; GENERAL INFORMATION:
; APPLICANT: Mitchell, William M.
; APPLICANT: Stratton, Charles W.
; TITLE OF INVENTION: DIAGNOSIS AND MANAGEMENT OF
; TITLE OF INVENTION: INFECTION CAUSED BY CHLAMYDIA
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA

```

COUNTRY: USA
 ZIP: 02173
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows 95
 SOFTWARE: FastSEQ for Windows Version 2.0b
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US 09/100,785
 FILING DATE: 19-Mar-2002
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/073,661
 FILING DATE: 06-MAY-1998
 APPLICATION NUMBER: US 09/025,521
 FILING DATE: 18-FEB-1998
 APPLICATION NUMBER: US 08/911,593
 FILING DATE: 14-AUG-1997
 APPLICATION NUMBER: US 60/023,921
 FILING DATE: 14-AUG-1996
 APPLICATION NUMBER: US 09/025,176
 FILING DATE: 18-FEB-1998
 APPLICATION NUMBER: US 09/025,174
 FILING DATE: 18-FEB-1998
 APPLICATION NUMBER: US 60/045,739
 FILING DATE: 6-MAY-1997
 APPLICATION NUMBER: US 60/045,779
 FILING DATE: 6-MAY-1997
 APPLICATION NUMBER: US 60/045,780
 FILING DATE: 6-MAY-1997
 APPLICATION NUMBER: US 60/045,784
 FILING DATE: 6-MAY-1997
 APPLICATION NUMBER: US 60/045,787
 FILING DATE: 6-MAY-1997
 APPLICATION NUMBER: US 60/045,689
 FILING DATE: 6-MAY-1997

ATTORNEY/AGENT INFORMATION:
 NAME: Carroll, Alice O
 REGISTRATION NUMBER: 33,542
 REFERENCE/DOCKET NUMBER: VDB97-09pm
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 781-861-6240
 TELEFAX: 781-861-9540
 TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 50:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 101 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 SEQUENCE DESCRIPTION: SEQ ID NO: 50:
 US-10-100-785-50

Query Match 31.2%; Score 5; DB 4; Length 101;
 Best Local Similarity 100.0%; Pred.No.1.3e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 NKEFK 14
 |||||
 Db 79 NKEFK 83

RESULT 15
 US-09-325-932A-51
 ; Sequence 51, Application US/09325932A
 ; Patent No. 6451604
 ; GENERAL INFORMATION:
 ; APPLICANT: Flinn, Barry
 ; APPLICANT: Lasham, Annette
 ; TITLE OF INVENTION: Compositions affecting programmed cell
 ; TITLE OF INVENTION: death and their use in the modification of forestry plant develop
 ; FILE REFERENCE: 1022

CURRENT APPLICATION NUMBER: US/09/325,932A
 CURRENT FILING DATE: 1999-06-04
 NUMBER OF SEQ ID NOS: 206
 SOFTWARE: FastSEQ for Windows Version 3.0
 SEQ ID NO 51
 LENGTH: 115
 TYPE: PRT
 ORGANISM: Pinus radiata
 US-09-325-932A-51

Query Match 31.2%; Score 5; DB 4; Length 115;
 Best Local Similarity 100.0%; Pred.No.1.5e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 NKEFK 14
 |||||
 Db 84 NKEFK 88

Search completed: April 23, 2004, 14:55:49
 Job time : 15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:54:17 ; Search time 35.6667 Seconds
(without alignments)
124.026 Million cell updates/sec

Title: US-09-528-682-3_COPY_64_79

Perfect score: 16

Sequence: 1 MSTFEQVPNNKEFKGV 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1133595 seqs, 276475211 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1133595

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09D_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US09E_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	50.0	241	9	US-09-819-917-5
2	6	37.5	51	12	US-10-424-599-212966
3	6	37.5	66	12	US-10-424-599-177635
4	6	37.5	77	14	US-10-106-698-5318
5	6	37.5	124	9	US-09-864-761-3351
6	6	37.5	289	9	US-09-765-272-20
7	6	37.5	296	12	US-10-282-1222A-52171
8	6	37.5	309	10	US-09-769-787-156
9	6	37.5	309	10	US-09-754-809-2
10	6	37.5	309	12	US-10-455-109-2
11	6	37.5	319	15	US-10-585-198-174
12	6	37.5	440	12	US-10-282-1222A-48298
13	6	37.5	476	14	US-10-316-253-157
14	6	37.5	476	14	US-10-316-253-159
15	6	37.5	512	10	US-09-976-800-102

16	6	37.5	512	10	US-09-976-800-103	Sequence 103, App
17	6	37.5	512	10	US-09-976-800-114	Sequence 114, App
18	6	37.5	512	10	US-09-976-800-115	Sequence 115, App
19	6	37.5	512	14	US-10-138-838-102	Sequence 102, App
20	6	37.5	512	14	US-10-138-838-103	Sequence 103, App
21	6	37.5	512	14	US-10-138-838-114	Sequence 114, App
22	6	37.5	512	14	US-10-138-838-115	Sequence 115, App
23	6	37.5	512	14	US-10-139-031-102	Sequence 102, App
24	6	37.5	512	14	US-10-139-031-103	Sequence 103, App
25	6	37.5	512	14	US-10-139-031-114	Sequence 114, App
26	6	37.5	512	14	US-10-139-031-115	Sequence 115, App
27	6	37.5	512	14	US-10-138-905-102	Sequence 102, App
28	6	37.5	512	14	US-10-138-905-103	Sequence 103, App
29	6	37.5	512	14	US-10-138-905-114	Sequence 114, App
30	6	37.5	512	14	US-10-138-905-115	Sequence 115, App
31	6	37.5	512	14	US-10-139-296-102	Sequence 102, App
32	6	37.5	512	14	US-10-139-296-103	Sequence 103, App
33	6	37.5	512	14	US-10-139-296-114	Sequence 114, App
34	6	37.5	512	14	US-10-139-296-115	Sequence 115, App
35	6	37.5	512	14	US-10-139-218-102	Sequence 102, App
36	6	37.5	512	14	US-10-139-218-103	Sequence 103, App
37	6	37.5	512	14	US-10-139-218-114	Sequence 114, App
38	6	37.5	512	14	US-10-139-218-115	Sequence 115, App
39	6	37.5	512	14	US-10-405-660-102	Sequence 102, App
40	6	37.5	512	14	US-10-405-660-103	Sequence 103, App
41	6	37.5	512	14	US-10-405-660-114	Sequence 114, App
42	6	37.5	512	14	US-10-405-660-115	Sequence 115, App
43	6	37.5	512	14	US-10-405-660-102	Sequence 102, App
44	6	37.5	512	14	US-10-405-660-103	Sequence 103, App
45	6	37.5	512	14	US-10-405-660-114	Sequence 114, App

ALIGNMENTS

RESULT 1
US-09-819-917-5
; Sequence 5, Application US/09819917
; Patent No. US20020044939A1
; GENERAL INFORMATION:
; APPLICANT: Fizza, Mariagrazia
; APPLICANT: Fontana, Maria Rita
; APPLICANT: Giannelli, Valentina
; APPLICANT: Rapuoli, Rina
; TITLE OF INVENTION: Immunogenic Detoxified Mutants Of Cholera Toxin
; FILE REFERENCE: CHIRO312
; CURRENT APPLICATION NUMBER: US/09/819,917
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: 08/981,208
; PRIOR FILING DATE: 1997-12-22
; PRIOR APPLICATION NUMBER: 9513371.6
; PRIOR FILING DATE: 1995-06-30
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 5
; LENGTH: 241
; TYPE: PRT
; ORGANISM: E. coli
US-09-819-917-5

Query Match 50.0%; Score 8; DB 9; Length 241;
Best Local Similarity 100.0%; Pred. No. 0.42; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 VPNNKEFK 14
Db 186 VPNNKEFK 193

RESULT 2
US-10-424-599-212966
; Sequence 212966, Application US/10424599
; Publication No. US20040031072A1

; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 212966
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_34333C.1.pep
US-10-424-599-212966

Query Match 37.5%; Score 6; DB 12; Length 51;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 KEFKGV 16
Db 22 KEFKGV 27

RESULT 3
US-10-424-599-177635
; Sequence 177635, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 177635
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(66)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_13141C.1.pep
US-10-424-599-177635

Query Match 37.5%; Score 6; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PNNKEF 13
Db 22 PNNKEF 27

RESULT 4
US-10-106-698-5318
; Sequence 5318, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698

; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5318
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5318

Query Match 37.5%; Score 6; DB 14; Length 77;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 VPNNKE 12
Db 54 VPNNKE 59

RESULT 5
US-09-864-761-33551
; Sequence 33551, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203

;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 33551
;; LENGTH: 124
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL109618.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1
;; OTHER INFORMATION: EXPRESSED IN HELL100, SIGNAL = 16
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.5
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 7.5
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.7
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.9
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.5
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9.6
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 23
;; OTHER INFORMATION: EST HUMAN HIT: BE22562.1, EVALUATE 2.00e-30
;; OTHER INFORMATION: EST_HUMAN HIT: A1625087.1, EVALUATE 2.00e-30
US-09-864-761-33551

Query Match 37.5%; Score 6; DB 9; Length 124;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 VPNNKE 12
|||||
Db 101 VPNNKE 106

RESULT 6
US-09-765-272-20
; Sequence 20, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 289 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

;; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-765-272-20

Query Match 37.5%; Score 6; DB 9; Length 289;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PNNKEF 13
|||||
Db 143 PNNKEF 148

RESULT 7
US-10-282-122A-52171
; Sequence 52171, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52171
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-52171

Query Match 37.5%; Score 6; DB 12; Length 296;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 NKEFKG 15
|||||
Db 69 NKEFKG 74

RESULT 8
US-09-769-787-156

Sequence 156, Application US/09769787
 Publication No. US20030091577A1
 GENERAL INFORMATION:
 APPLICANT: Microbial Technics Limited
 APPLICANT: Gilbert, Christophe FG
 APPLICANT: Hansbro, Philip M
 TITLE OF INVENTION: Proteins
 FILE REFERENCE: PWC/P21129W0
 CURRENT APPLICATION NUMBER: US/09/769,787
 PRIOR FILING DATE: 2001-01-26
 PRIOR APPLICATION NUMBER: GB 9816337.1
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: US 60/125164
 PRIOR FILING DATE: 1999-03-19
 NUMBER OF SEQ ID NOS: 388
 SOFTWARE: Patent In Ver. 2.1
 SEQ ID NO 156
 LENGTH: 309
 TYPE: PRT
 ORGANISM: Streptococcus pneumoniae
 US-09-769-787-156

Query Match 37.5%; Score 6; DB 10; Length 309;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PNNKEF 13
 |||||
 DB 163 PNNKEF 168

RESULT 9

US-09-754-809-2
 Sequence 2, Application US/09754809
 Publication No. US20030105307A1
 GENERAL INFORMATION:
 APPLICANT: SAMPSON, JACQUELYN S.
 APPLICANT: RUSSELL, HAROLD
 APPLICANT: THARPE, JEAN A.
 APPLICANT: ADES, EDWIN W.
 APPLICANT: CARLONE, GEORGE M.
 TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE 37 kDa SURFACE
 FILE REFERENCE: 64778 US
 CURRENT APPLICATION NUMBER: US/09/754,809
 PRIOR FILING DATE: 2001-01-03
 PRIOR APPLICATION NUMBER: US/09/221,753
 PRIOR FILING DATE: 1998-12-28
 PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 07/791,377
 PRIOR FILING DATE: EARLIER FILING DATE: 1991-09-17
 PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 07/816,286
 PRIOR FILING DATE: EARLIER FILING DATE: 1992-01-03
 PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 08/222,179
 PRIOR FILING DATE: EARLIER FILING DATE: 1994-04-04
 PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 08/715,131
 NUMBER OF SEQ ID NOS: 4
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 2
 LENGTH: 309
 TYPE: PRT
 ORGANISM: STREPTOCOCCUS PNEUMONIAE
 US-09-754-809-2

Query Match 37.5%; Score 6; DB 10; Length 309;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PNNKEF 13
 |||||
 DB 163 PNNKEF 168

RESULT 10
 US-10-455-109-2
 Sequence 2, Application US/10455109
 Publication No. US20030204074A1
 GENERAL INFORMATION:
 APPLICANT: SAMPSON, JACQUELYN S.
 APPLICANT: RUSSELL, HAROLD
 APPLICANT: THARPE, JEAN A.
 APPLICANT: ADES, EDWIN W.
 APPLICANT: CARLONE, GEORGE M.
 TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE 37 kDa SURFACE
 FILE REFERENCE: 64778 US
 CURRENT APPLICATION NUMBER: US/10/455,109
 PRIOR FILING DATE: 2003-06-04
 PRIOR APPLICATION NUMBER: US/09/221,753
 PRIOR FILING DATE: 1998-12-28
 PRIOR APPLICATION NUMBER: US 07/791,377
 PRIOR FILING DATE: 1991-09-17
 PRIOR APPLICATION NUMBER: US 07/816,286
 PRIOR FILING DATE: 1992-01-03
 PRIOR APPLICATION NUMBER: US 08/222,179
 PRIOR FILING DATE: 1994-04-04
 PRIOR APPLICATION NUMBER: US 08/715,131
 PRIOR FILING DATE: 1996-09-17
 NUMBER OF SEQ ID NOS: 4
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 2
 LENGTH: 309
 TYPE: PRT
 ORGANISM: STREPTOCOCCUS PNEUMONIAE
 US-10-455-109-2

Query Match 37.5%; Score 6; DB 12; Length 309;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PNNKEF 13
 |||||
 DB 163 PNNKEF 168

RESULT 11

US-10-085-198-174
 Sequence 174, Application US/10085198
 Publication No. US20040009907A1
 GENERAL INFORMATION:
 APPLICANT: Alsbrook et al.
 TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
 FILE REFERENCE: 21402-279
 CURRENT APPLICATION NUMBER: US/10/085,198
 CURRENT FILING DATE: 2002-02-25
 PRIOR APPLICATION NUMBER: 60/271,646
 PRIOR FILING DATE: 2001-02-26
 PRIOR APPLICATION NUMBER: 60/276,401
 PRIOR FILING DATE: 2001-03-16
 PRIOR APPLICATION NUMBER: 60/311,981
 PRIOR FILING DATE: 2001-08-13
 PRIOR APPLICATION NUMBER: 60/312,858
 PRIOR FILING DATE: 2001-08-16
 PRIOR APPLICATION NUMBER: 60/271,840
 PRIOR FILING DATE: 2001-02-27
 PRIOR APPLICATION NUMBER: 60/277,324
 PRIOR FILING DATE: 2001-03-20
 PRIOR APPLICATION NUMBER: 60/286,096
 PRIOR FILING DATE: 2001-04-21
 PRIOR APPLICATION NUMBER: 60/299,695
 PRIOR FILING DATE: 2001-06-20
 PRIOR APPLICATION NUMBER: 60/315,614
 PRIOR FILING DATE: 2001-08-29
 PRIOR APPLICATION NUMBER: 60/272,405
 PRIOR FILING DATE: 2001-02-28
 Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 653
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 174
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-085-198-174

Query Match 37.5%; Score 6; DB 15; Length 319;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 NKEPKG 15
|||||
Db 296 NKEPKG 301

RESULT 12

US-10-282-122A-48298
; Sequence 48298, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITFA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 48298

; LENGTH: 440

; TYPE: PRT

; ORGANISM: Bacteroides fragilis

US-10-282-122A-48298

Query Match 37.5%; Score 6; DB 12; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 STFEQV 7
|||||

Db 389 STFEQV 394

RESULT 13

US-10-316-253-157
; Sequence 157, Application US/10316253
; Publication No. US20030162706A1
; GENERAL INFORMATION:
; APPLICANT: The Procter & Gamble Company
; APPLICANT: Peters, Kevin
; APPLICANT: Thompson, Larry
; APPLICANT: Wang, Feng
; APPLICANT: Greis, Kenneth
; TITLE OF INVENTION: Angiogenesis Modulating Proteins
; FILE REFERENCE: 8865M
; CURRENT APPLICATION NUMBER: US/10/316,253
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/355,295
; PRIOR FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 157
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-316-253-157

Query Match 37.5%; Score 6; DB 14; Length 476;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 STFEQV 7
|||||

Db 55 STFEQV 60

RESULT 14

US-10-316-253-159
; Sequence 159, Application US/10316253
; Publication No. US20030162706A1
; GENERAL INFORMATION:
; APPLICANT: The Procter & Gamble Company
; APPLICANT: Peters, Kevin
; APPLICANT: Thompson, Larry
; APPLICANT: Wang, Feng
; APPLICANT: Greis, Kenneth
; TITLE OF INVENTION: Angiogenesis Modulating Proteins
; FILE REFERENCE: 8865M
; CURRENT APPLICATION NUMBER: US/10/316,253
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/355,295
; PRIOR FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 159
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-316-253-159

Query Match 37.5%; Score 6; DB 14; Length 476;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 STFEQV 7
|||||

Db 55 STFEQV 60

RESULT 15

US-09-976-800-102
; Sequence 102, Application US/09976800
; Publication No. US20030077795A1

; GENERAL INFORMATION:
; APPLICANT: Wilson, Ron C.
; APPLICANT: Craft, David L.
; APPLICANT: Erlich, Dudley
; APPLICANT: Eshoo, Mark
; APPLICANT: Madduri, Krishna M.
; APPLICANT: Cornett, Cathy A.
; APPLICANT: Brenner, Alfred A.
; APPLICANT: Tang, Maria
; APPLICANT: Loper, John C.
; APPLICANT: Gleeson, Martin
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450 OXIDOREDUCTASE
; TITLE OF INVENTION: GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF C. ALBICOLA
; TITLE OF INVENTION: TROPICALIS AND METHODS RELATING THERETO
; FILE REFERENCE: 1010-16
; CURRENT APPLICATION NUMBER: US/09/976,800
; CURRENT FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 102
; LENGTH: 512
; TYPE: PRT
; ORGANISM: CANDIDATROPICALIS
US-09-976-800-102

Query Match 37.5%; Score 6; DB 10; Length 512;
Best Local Similarity 100.0%; Pred.No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 NNKEFK 14
|||||
Db 246 NNKEFK 251

Search completed: April 23, 2004, 15:07:24
Job time : 35.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:48:31 ; Search time 11.6667 Seconds
(without alignments)
131.920 Million cell updates/sec

Title: US-09-528-682-4_COPY_64_79
Perfect score: 16
Sequence: 1 VREFNSLPNNKASDST 16

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	43.8	105	2	C72631
2	6	37.5	97	2	T39890
3	6	37.5	124	2	S67606
4	6	37.5	141	2	G86688
5	6	37.5	191	2	A97092
6	6	37.5	289	2	S75361
7	6	37.5	380	2	C90119
8	6	37.5	399	2	AD2559
9	6	37.5	465	2	T27885
10	6	37.5	472	2	AF2803
11	6	37.5	472	2	F97582
12	6	37.5	502	2	T32041
13	6	37.5	634	2	F97172
14	6	37.5	689	2	S45901
15	6	37.5	699	2	JG6132
16	6	37.5	700	1	S12053
17	6	37.5	844	2	AC1981
18	6	37.5	887	2	S64016
19	6	37.5	1077	2	T41146
20	5	31.2	5	2	PT0561
21	5	31.2	6	2	PT0619
22	5	31.2	44	2	B39137
23	5	31.2	51	2	T07344
24	5	31.2	69	2	C69790
25	5	31.2	69	2	C96021
26	5	31.2	77	2	S30873
27	5	31.2	77	2	AC3608
28	5	31.2	84	2	T03787
29	5	31.2	86	2	S66089

30	5	31.2	88	2	A03317
31	5	31.2	95	2	AG0364
32	5	31.2	102	2	D71045
33	5	31.2	104	2	I47165
34	5	31.2	104	2	S45400
35	5	31.2	106	2	S69868
36	5	31.2	109	2	S29387
37	5	31.2	112	2	S45968
38	5	31.2	117	2	H90021
39	5	31.2	125	2	C70162
40	5	31.2	125	2	D72247
41	5	31.2	126	2	C70307
42	5	31.2	142	2	AD0923
43	5	31.2	143	2	H84456
44	5	31.2	144	2	I51047
45	5	31.2	147	2	AH2287

ALIGNMENTS

RESULT 1
C72631
hypothetical protein APE1506 - Aeropyrum pernix (strain KI)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: C72631
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: C72631
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-105 <KAW>
A:Cross-references: DDBJ:AP000061; NID:G5104821; PIDN:BAA80505.1; PID:dl044291; PID:G510 A:Experimental source: strain KI
C:Genetics:
A:Gene: APE1506

Query Match 43.8%; Score 7; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LPNNKAS 13
|||||
DB 57 LPNNKAS 63

RESULT 2

T39890
hypothetical protein SPBC21.08c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39890
R;Saunders, D.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G. submitted to the EMBL Data Library, February 1999
A:Reference number: Z21860
A:Accession: T39890
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-97 <SAU>
A:Cross-references: ENML:AL035537; PIDN:CNB36888.1; GSPDB:GN00067; SPDB:SPBC21.08c A:Experimental source: strain 972h-; cosmid c21
C:Genetics:
A:Map position: 2
A:Introns: 5/1

Query Match 37.5%; Score 6; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      5 NSLPNN 10
        |||||
Db      67 NSLPNN 72

RESULT 3
S67606
Proble membrane protein YDL071c - Yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein D2502
C:Species: Saccharomyces cerevisiae
C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 19-Apr-2002
C:Accession: S67606
R:Blöcker, H.; Brandt, P.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67587
A:Accession: S67606
A:Molecule type: DNA
A:Residues: 1-124 <BLO>
A:Cross-references: EMBL:Z74119; NID:G1431080; PID:G1431081; GSPDB:GN00004; MIPS:YDL071c
A:Experimental source: strain S288C
C:Genetics:
A:Gene: MIPS:YDL071c
A:Cross-references: SGD:S0002229
A:Map position: 4L
C:Superfamily: Saccharomyces probable membrane protein YDL071c
C:Keywords: transmembrane protein
F:187-103/Domain: transmembrane #status predicted <TM>

Query Match      37.5%; Score 6; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 SLPNNK 11
        |||||
Db      40 SLPNNK 45

RESULT 4
G86688
Prophage p2 protein 09 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: G86688
R:Bototin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: G86688
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-141 <STO>
A:Cross-references: GB:AF005176; PID:G12723394; PIDN:AAK04609.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: ps209

Query Match      37.5%; Score 6; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 REFNSL 7
        |||||
Db     114 REFNSL 119

RESULT 5
A97092
probable acetyltransferase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: A97092
R:Nolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

```

```

; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clc
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: A97092
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-191 <KUR>
A:Cross-references: GB:AB001437; PIDN:AAK79524.1; PID:G15024508; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1557

```

```

Query Match      37.5%; Score 6; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      8 PNNKAS 13
        |||||
Db     135 PNNKAS 140

```

RESULT 6

```
S75361
dihydropteroate synthase (EC 2.5.1.15) - Synecocystis sp. (strain PCC 6803)

```

```
N:Alternate names: dihydropteroate pyrophosphorylase; protein slr2026
C:Species: Synecocystis sp.
A:Variety: PCC 6803

```

```
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 02-Sep-2000
C:Accession: S75361

```

```
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis

```

```
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231

```

```
A:Accession: S75361

```

```
A>Status: nucleic acid sequence not shown; translation not shown

```

```
A:Molecule type: DNA

```

```
A:Residues: 1-289 <KAN>

```

```
A:Cross-references: EMBL:D90904; GB:AB001339; NID:G1522225; PIDN:BAAL17275.1; PID:G152235

```

```
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:

```

```
A:Gene: folp

```

```
C:Superfamily: dihydropteroate synthase; dihydropteroate synthase homology

```

```
C:Keywords: folate biosynthesis; transferase

```

```
F:30-272/Domain: dihydropteroate synthase homology <DHS>

```

```
Query Match      37.5%; Score 6; DB 2; Length 289;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
QY      3 EFNSLP 8
        |||||
Db     46 EFNSLP 51

```

RESULT 7

```
C90119
hypothetical protein orf380 [imported] - Guillardia theta nucleomorph

```

```
C:Species: nucleomorph Guillardia theta

```

```
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont

```

```
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: C90119

```

```
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei
Nature 410, 1091-1096, 2001

```

```
A:Title: The highly reduced genome of an enslaved algal nucleus.

```

```
A:Reference number: A99082; MUID:11323671; PMID:11323671

```

```
A:Accession: C90119

```

```
A>Status: preliminary

```

```
A:Molecule type: DNA

```

```
A:Residues: 1-380 <DOU>

```

```
A:Cross-references: GB:AF083031; NID:G13794272; PIDN:AAK39649.1; GSPDB:GN00152

```

C;Genetics:

A;Gene: orf380
A;Map position: 3
A;Genome: nucleomorph
C;Keywords: nucleomorph

Query Match 37.5%; Score 6; DB 2; Length 380;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 NNKASS 14

|||||
Db 97 NNKASS 102

RESULT 8

AD2559
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AD2559
R;Kaneoko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AD2559
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-399 <KUR>
A;Cross-references: GB:AP003603; PIDN:BAW77397.1; PID:gi71134840; GSPDB:GN00182
C;Genetics:
A;Gene: all8067
A;Genome: plasmid

Query Match 37.5%; Score 6; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 NNKASS 14

|||||
Db 88 NNKASS 93

RESULT 9

T27885
sperm membrane protein spe-4 - Caenorhabditis elegans
N;Alternate names: probable integral membrane protein
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Nov-2002
C;Accession: T27885; A43459; S24632; S24633
R;Gardner, A.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z20435
A;Accession: T27885
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-465 <WIL>
A;Cross-references: EMBL:Z73912; PIDN:CAA98145.1; GSPDB:GN00019; CESP:ZK524.1
R;J. Hernault, S.W.; Arduengo, P.M.
J. Cell Biol. 119, 55-68, 1992
A;Title: Mutation of a putative sperm membrane protein in Caenorhabditis elegans prevents the formation of a functional sperm
A;Reference number: A43459; MUID:92407040; PMID:1527173
A;Accession: A43459
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-465 <LHE>
A;Cross-references: EMBL:Z14067; NID:96868; PID:96870; EMBL:Z14066; NID:96870; PID:96871
A;Experimental source: strain Bristol N2
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1992

C;Genetics:

A;Gene: CESP:ZK524.1
A;Map position: 1
A;Introns: 69/3; 154/3; 200/1; 224/3; 300/1; 386/1; 435/1
C;Superfamily: presenilin

Query Match 37.5%; Score 6; DB 2; Length 465;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NSLPNN 10

|||||
Db 366 NSLPNN 371

RESULT 10

AF2803
transcription regulator, GntR family Atu1848 [imported] - Agrobacterium tumefaciens (str C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AF2803
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; Karp, P.; Grant, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McClellan, S.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AF2803
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-472 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL42844.1; PID:gi7740293; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu1848
A;Map position: circular chromosome

Query Match 37.5%; Score 6; DB 2; Length 472;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8

|||||
Db 385 EFNSLP 390

RESULT 11

F97582
probable transcription regulator (PA2100) [imported] - Agrobacterium tumefaciens (strain C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: F97582
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourullo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: F97582
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-472 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87615.1; PID:gl5156961; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_3389
A;Map position: circular chromosome

Query Match 37.5%; Score 6; DB 2; Length 472;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8

Db 385 EFNSLP 390
|||||

RESULT 12

T32041

hypothetical protein F41B5.9 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C;Accession: T32041

R;Dante, M.; Kramer, J.

submitted to the EMBL Data Library, July 1997

A;Description: The sequence of C. elegans cosmid F41B5.

A;Reference number: Z21115

A;Accession: T32041

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-502 <DAN>

A;Cross-references: EMBL:AF016676; PIDN:AAC25905.1; GSPDB:GN00023; CESP:F41B5.9

A;Experimental source: strain Bristol N2; clone F41B5

C;Genetics:

A;Gene: CESP:F41B5.9

A;Map position: 5

A;Introns: 6/3; 24/3; 59/2; 106/1; 310/2; 429/2; 451/3

Query Match

Best Local Similarity 37.5%; Score 6; DB 2; Length 502;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8

Db 236 EFNSLP 241

RESULT 13

F97172

flagellar hook-associated protein FlgK [imported] - Clostridium acetobutylicum

C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C;Accession: F97172

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 193, 4823-4838, 2001

A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld

A;Reference number: A96900; MUID:21359325; PMID:21359325

A;Accession: F97172

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-634 <KUR>

A;Cross-references: GB:AE001437; PIDN:AAK80169.1; PID:gl5025209; GSPDB:GN00168

A;Experimental source: Clostridium acetobutylicum ATCC824

C;Genetics:

A;Gene: CAC2212

Query Match

Best Local Similarity 37.5%; Score 6; DB 2; Length 634;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 NKASSD 15

Db 297 NKASSD 302

RESULT 14

S45901

probable membrane protein YBR043C - yeast (Saccharomyces cerevisiae)

N;Alternate names: hypothetical protein YBR0413

C;Species: Saccharomyces cerevisiae

C;Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 19-Apr-2002

C;Accession: S45901

R;Andre, B.; Cziepluch, C.; Hein, C.; Jauniaux, J.C.; Urrestarazu, A.; Vissers, S.

submitted to the Protein Sequence Database, August 1994

A;Reference number: S45893

A;Accession: S45901
A;Molecule type: DNA
A;Residues: 1-689 <AND>
A;Cross-references: EMBL:Z35912; NID:G536267; PIDN:CAA84985.1; PID:G536268; GSPDB:GN0000
A;Experimental source: strain S288C
C;Genetics:

A;Gene: MIPS:YBR043C

A;Cross-references: SGD:S0000247

A;Map position: 2R

C;Superfamily: Yeast probable membrane protein YBR043C

C;Keywords: transmembrane protein

F;108-124/Domain: transmembrane #status predicted <TM1>

F;140-161/Domain: transmembrane #status predicted <TM2>

F;177-193/Domain: transmembrane #status predicted <TM3>

F;239-256/Domain: transmembrane #status predicted <TM4>

F;285-287/Domain: transmembrane #status predicted <TM5>

F;476-493/Domain: transmembrane #status predicted <TM6>

F;513-529/Domain: transmembrane #status predicted <TM7>

F;555-579/Domain: transmembrane #status predicted <TM8>

F;587-603/Domain: transmembrane #status predicted <TM9>

F;629-645/Domain: transmembrane #status predicted <TM10>

F;648-670/Domain: transmembrane #status predicted <TM11>

Query Match

Best Local Similarity 37.5%; Score 6; DB 2; Length 689;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NSLPNN 10

Db 20 NSLPNN 25

RESULT 15

JC6132

protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type epsilon precursor - mouse

N;Alternate names: phosphotyrosine phosphatase; protein-tyrosine-phosphate phosphohydrol

C;Species: Mus musculus (house mouse)

C;Date: 16-Apr-1997 #sequence_revision 09-May-1997 #text_change 21-Jun-2002

C;Accession: JC6132

R;Schmidt, A.; Rutledge, S.J.; Endo, N.; Opas, E.E.; Tanaka, H.; Wesolowski, G.; Liu, C.

Proc. Natl. Acad. Sci. U.S.A. 93, 3068-3073, 1996

A;Title: Protein-tyrosine phosphatase activity regulates osteoclast formation and functi

A;Reference number: JC6132; MUID:96181534; PMID:8610169

A;Contents: bone marrow cell

A;Accession: JC6132

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-699 <SCH>

A;Cross-references: GB:U40280; NID:gl373052; PIDN:AAB02190.1; PID:gl373053

C;Comment: This enzyme plays an important role in osteoclast formation and function in t

hosphonate action.

C;Genetics:

A;Gene: ptp

C;Superfamily: protein-tyrosine-phosphatase, receptor type alpha; leukocyte common antig

C;Keywords: phosphoprotein; phosphoric monoester hydrolase; transmembrane protein; tyros

F;77-697/Domain: leukocyte common antigen cytosolic domain homology <LAC>

F;158-382/Domain: protein-tyrosine-phosphatase homology <PTPL>

F;334/Active site: Cys (phosphocysteine intermediate) #status predicted

F;340/Binding site: substrate phosphate (Arg) #status predicted

F;629/Active site: Cys (phosphocysteine intermediate) #status predicted

F;635/Binding site: substrate phosphate (Arg) #status predicted

Query Match

Best Local Similarity 37.5%; Score 6; DB 2; Length 699;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8

Db 137 EFNSLP 142

Search completed: April 23, 2004, 14:54:57

Job time : 12.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:47:06 ; Search time 8 Seconds
(without alignments)
104.140 Million cell updates/sec

Title: US-09-528-682-4_COPY_64_79
Perfect score: 16
Sequence: 1 VREFNSLPNNKASDST 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	62.5	263	1 E2BA_ECOLI	P43528 escherichia
2	6	37.5	289	1 DHPS_SNY3	P73248 synecocyst
3	6	37.5	350	1 DPO4_FUSNN	Q8reb0 fusobacteri
4	6	37.5	428	1 SPN7_SCHPO	O60165 schizosacch
5	6	37.5	449	1 BPL3_MOUSE	Q8bu51 mus musculu
6	6	37.5	465	1 SP24_CAEEL	Q01608 caenorhabdi
7	6	37.5	596	1 GLMN_MOUSE	Q8bzml mus musculu
8	6	37.5	689	1 YBP3_YEAST	P38227 saccharomyc
9	6	37.5	699	1 PTPE_MOUSE	P49446 mus musculu
10	6	37.5	700	1 PTPE_HUMAN	P23469 homo sapien
11	6	37.5	888	1 PUF4_YEAST	P25339 saccharomyc
12	6	37.5	1162	1 ITAD_HUMAN	Q13349 homo sapien
13	5	31.2	42	1 GB57_MOUSE	Q61016 mus musculu
14	5	31.2	42	1 LPAS_STRMU	P10540 streptococc
15	5	31.2	66	1 RGSA_RAT	P49806 rattus norv
16	5	31.2	84	1 MT22_ORYSA	P93433 oryza sativ
17	5	31.2	86	1 YBKO_BACSU	P37557 bacillus su
18	5	31.2	104	1 YBKO_YEAST	P38168 saccharomyc
19	5	31.2	109	1 NIFW_AZOCA	P26481 azorhizobiu
20	5	31.2	112	1 YBVO_YEAST	P38259 saccharomyc
21	5	31.2	125	1 RS13_BORBU	O51453 borrelia bu
22	5	31.2	125	1 RS13_THEMA	Q9x15 thermotoga
23	5	31.2	126	1 RS13_AQUAE	O66486 aquifex aeo
24	5	31.2	144	1 LYC2_ONCMY	P11941 oncorhynchu
25	5	31.2	153	1 VE6_MMPV	P30734 micromys mi
26	5	31.2	155	1 IGF2_BOVIN	P07456 bos taurus
27	5	31.2	167	1 YLXL_BACSU	P40405 bacillus su
28	5	31.2	178	1 CAVI_BOVIN	P79132 bos taurus
29	5	31.2	178	1 CAVI_CANFA	P33724 canis fami
30	5	31.2	178	1 CAVI_HUMAN	Q03135 homo sapien
31	5	31.2	178	1 CAVI_MOUSE	P49817 mus musculu
32	5	31.2	178	1 CAVI_RAT	P41350 rattus norv
33	5	31.2	179	1 UCRI_RHOVI	P81380 rhodopseudo

RESULT 1
E2BA_ECOLI
ID E2BA_ECOLI STANDARD; PRT; 263 AA.
AC P43528;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heat-labile enterotoxin IIB, A chain precursor (LT-IIB).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate 41;
RX MEDLINE=89359131; PubMed=2670900;
RA Pickett C.L., Twiddy E.M., Coker C., Holmes R.K.;
RT "Cloning, nucleotide sequence, and hybridization studies of the type
IIB heat-labile enterotoxin gene of Escherichia coli.";
RL J. Bacteriol. 171:4945-4952 (1989).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS).
RX MEDLINE=96399713; PubMed=8805549;
RA van den Akker F., Sarfaty S., Twiddy E.M., Connell T.D., Holmes R.K.,
RA Hol W.G.J.;
RT "Crystal structure of a new heat-labile enterotoxin, LT-IIB.";
Structure 4:665-678 (1996).
CC -!- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC -!- SUBUNIT: HETEROHXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M28523; AAA53285.1; -.
CC PDB; 1TII; 17-AUG-96.
DR InterPro: IPR001144; Enterotoxin_A.
DR Pfam: PF01375; Enterotoxin_A; 1.
DR PRINTS; PR00771; ENTEROTOXINA.
KW Enterotoxin; Signal; 3D-structure.
FT SIGNAL 1 20
FT CHAIN 21 263 HEAT-LABILE ENTEROTOXIN IIB, A CHAIN.
FT DISULFID 205 217
FT ACT SITE 130 130
FT STRAND 22 27
FT HELIX 31 37
FT TURN 38 38
FT STRAND 39 40
FT TURN 43 44
FT TURN 49 51
FT HELIX 59 64

34 5 31.2 181 1 RGSA_MOUSE
35 5 31.2 182 1 LEUD_MYCSM
36 5 31.2 183 1 ATPD_ANTSP
37 5 31.2 198 1 RECR_CLOAB
38 5 31.2 204 1 PTMT_RHIME
39 5 31.2 209 1 HXA7_XENLA
40 5 31.2 220 1 HB7A_XENLA
41 5 31.2 220 1 UNG_BORBU
42 5 31.2 223 1 V053_FOWPV
43 5 31.2 225 1 R2_GALSU
44 5 31.2 233 1 YCA2_PLAFA
45 5 31.2 241 1

ALIGNMENTS

Q9cqe5 mus musculu
Q938c8 mycobacteri
Q02849 antithamio
Q97mr4 clostridium
Q08249 rhizobium m
P09071 xenopus lae
P04771 xenopus lae
P04476 xenopus lae
O51082 borrelia bu
P32817 fowlpox vir
P35014 galdieria s
P25408 plasmodium

```

FT TURN 75 76
FT STRAND 77 78
FT STRAND 81 81
FT HELIX 84 94
FT TURN 96 97
FT STRAND 100 107
FT TURN 110 111
FT STRAND 112 114
FT HELIX 115 119
FT STRAND 120 122
FT HELIX 126 128
FT STRAND 131 134
FT TURN 135 135
FT STRAND 137 138
FT HELIX 139 141
FT STRAND 142 149
FT TURN 150 151
FT STRAND 152 153
FT STRAND 158 159
FT TURN 161 162
FT HELIX 165 168
FT TURN 169 170
FT STRAND 174 174
FT HELIX 176 180
FT TURN 181 182
FT TURN 187 188
FT HELIX 190 193
FT TURN 195 196
FT HELIX 197 199
FT TURN 200 200
FT TURN 203 205
FT HELIX 217 248
FT SEQUENCE 263 AA; 29485 MW; A56E8A069427CFB9 CRC64;

Query Match 62.5%; Score 10; DB 1; Length 263;
Best Local Similarity 100.0%; Pred. No. 0.00057;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LPNKKASDST 16
Db |||||
206 LPNKKASDST 215

```

```

RESULT 2
DHPS_SVNV3 STANDARD; PRT; 289 AA.
AC P73248;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Dihydropteroate synthase (EC 2.5.1.15) (DHPS) (Dihydropteroate
DE pyrophosphorylase).
GN FOLP OR SIR2026.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugitani M., Sasamoto S., Kimura T.,
RA Hoshino T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
CC -1- FUNCTION: DHPS catalyzes the formation of the immediate precursor
CC of folic acid. It is implicated in resistance to sulfonamide (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: 2-amino-4-hydroxy-6-hydroxymethyl-7,8-
CC dihydropteridine diphosphate + 4-aminobenzoate = diphosphate +

```

```

CC dihydropteroate.
CC -1- PATHWAY: Dihydrofolate biosynthesis; second step.
CC -1- SIMILARITY: Belongs to the DHPS family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90904; BAA17275.1; -.
CC PIR; S75361; S75361.
CC HSP; P26282; 1AJ2.
CC InterPro; IPR000489; Dhdropt_synth.
CC InterPro; IPR006390; DHPS.
CC Pfam; PF00809; Pterin_bind; 1.
CC TIGRFAMs; TIGR01496; DHPS_1; 1.
CC PROSITE; PS00792; DHPS_1; 1.
CC PROSITE; PS00793; DHPS_2; 1.
CC Antibiotic resistance; Transferase; Folate biosynthesis;
CC Complete proteome.
CC SEQUENCE 289 AA; 31602 MW; 76D28F8ACCEB9BA5 CRC64;

Query Match 37.5%; Score 6; DB 1; Length 289;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8
Db |||||
46 EFNSLP 51

RESULT 3
DHPS_FUSNN STANDARD; PRT; 350 AA.
ID DPO4_FUSNN
AC Q8REB0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase IV (EC 2.7.7.7) (Pol IV).
GN DINB OR FN1199.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyripides N., Overbeek R.;
RA "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
CC -1- FUNCTION: Poorly processive, error-prone DNA polymerase involved
CC in untargeted mutagenesis. Copies undamaged DNA at stalled
CC replication forks, which arise in vivo from mismatched or
CC misaligned primer ends. These misaligned primers can be extended
CC by polIV. Exhibits no 3'-5' exonuclease (proofreading) activity.
CC May be involved in translesional synthesis, in conjunction with
CC the beta clamp from polIII (By similarity).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).
CC -1- COFACTOR: Binds 2 magnesium ions per subunit (By similarity).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the DNA polymerase type-Y family.
CC -1- SIMILARITY: Contains 1 umuC domain.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

DR EMBL; AE010625; AAL95395.1; -;
 DR HAWAP; MF 01113; -; 1.
 DR InterPro; IPR001126; UMC _like.
 DR Pfam; PF00817; IMS; 1.
 DR PROSITE; PS0173; UMC; 1.
 KW Transferase; DNA-directed DNA polymerase; DNA replication; DNA repair;
 KW DNA-binding; Mutator protein; Magnesium; Complete proteome.
 FT DOMAIN 5 181 UMC.
 FT SITE 14 14 SUBSTRATE DISCRIMINATION (BY SIMILARITY).
 FT ACT SITE 100 100 BY SIMILARITY.
 FT METAL 9 9 MAGNESIUM (BY SIMILARITY).
 FT METAL 99 99 MAGNESIUM (BY SIMILARITY).
 SQ SEQUENCE 350 AA; 41056 MW; 630DD689F5C6228A CRC64;

Query Match 37.5%; Score 6; DB 1; Length 350;
 Best Local Similarity 100.0%; Pred.No. 17;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 REFNSL 7
 DB 258 REFNSL 263

RESULT 4
 ID SPN7_SCHPO STANDARD; PRT; 428 AA.
 AC O60165; O94648; Q96057;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Septin homolog spn7.
 GN SPN7 OR SPBC19F8.01C OR SPBC21.08C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21849401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leharz H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas R., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potaashkin J.,

RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880 (2002).
 RN [2]
 RP SEQUENCE OF 298-428 FROM N.A.
 RA Wu J.-Q., Pringle J.R.;
 RT "Roles of septins in the fission yeast S. pombe.";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION.
 RA Wu J.-Q., Pringle J.R.;
 RL Submitted (MAR-2001) to Swiss-Prot.
 CC -!- FUNCTION: SPORULATION SPECIFIC PROTEIN.
 CC -!- SUBCELLULAR LOCATION: CORTICES OF DEVELOPING SPORES.
 CC -!- SIMILARITY: Belongs to the septin family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

DR EMBL; AL023594; CAAL19121.1; -;
 DR EMBL; AL035537; CAB36888.1; -;
 DR EMBL; AF417166; AAL13302.1; -;
 DR GeneDB Spombe; SPBC19F8.01c; -;
 DR InterPro; IPR000038; GTP_Cell_Div.
 DR Pfam; PF00735; GTP_CDC; 1.
 DR ProDom; PD002565; GTP_Cell_Div; 1.
 KW Cell division; Sporulation; GTP-binding.
 FT NP_BIND 37 44
 FT SEQUENCE 428 AA; 49298 MW; CD0B4E1D82DF41A4 CRC64;

Query Match 37.5%; Score 6; DB 1; Length 428;
 Best Local Similarity 100.0%; Pred.No. 21;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NSLENN 10
 DB 398 NSLENN 403

RESULT 5
 BPL3_MOUSE
 ID BPL3_MOUSE STANDARD; PRT; 449 AA.
 AC Q8BU51; Q80Z08;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Bactericidal/permeability-increasing protein-like 3 precursor.
 GN BPII3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Olfactory epithelium;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feilings F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Bhat N.K.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Max S.I., Wang J., Haieh F.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Hong L.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Prange C.,
 RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.H.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RP SEQUENCE OF 335-449 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Ovary;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Niki K., Tomaru Y., Hasegawa Y., Nogi A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrester A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kani A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perle G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempie C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh N., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- SIMILARITY: Belongs to the BPI/LBP/Piunc superfamily. BPI/LBP
CC family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AK087711; BAC39978.1; ALT_INIT.
CC EMBL; BC048083; AAH48083.1; -.
CC InterPro; IPR001124; LBP_BPI_CETP.
CC SMART; SM00329; BPI2; 1.
CC PROSITE; PS00400; LBP_BPI_CETP; FALSE_NEG.
CC -----
CC SIGNAL 1 18 POTENTIAL.
CC CHAIN 19 449 BACTERICIDAL/PERMEABILITY-INCREASING
CC PROTEIN-LIKE 3.
CC FT CARBOHYD 115 115 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CONFLICT 385 386 FF -> SS (IN REF. 2).
CC FT SEQUENCE 449 AA; 48901 MW; EF4290C3C9301DD6 CRC64;
CC -----
CC Query Match 37.5%; Score 6; DB 1; Length 449;
CC Best Local Similarity 100.0%; Pred. No. 21;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 7 LPNKA 12
CC Db 149 LPNKA 154

RESULT 6
SPEC CAEL
ID SPE4 CAEL STANDARD; PRT; 465 AA.
AC Q01608;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Presenilin spe-4.
GN SPE-4 OR ZK524.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoidea; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=92407040; PubMed=1527173;
RA L'Hernault S.W., Arduengo P.M.;
RT "Mutation of a putative sperm membrane protein in Caenorhabditis
RT elegans prevents sperm differentiation but not its associated meiotic
RT divisions.";
RL J. Cell Biol. 119:55-68 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Gardner A.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP FUNCTION, SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND MUTANTS HC78
RP AND EB12.
RA MEDLINE=99038173; PubMed=9819355;
RA Arduengo P.M., Appleberry O.K., Chuang P., L'Hernault S.W.;
RT "The presenilin protein family member SPE-4 localizes to an ER/Golgi
RT derived organelle and is required for proper cytoplasmic partitioning
RT during Caenorhabditis elegans spermatogenesis.";
RT J. Cell Sci. 111:3645-3654 (1998).
CC -!- FUNCTION: Potential catalytic subunit of the gamma-secretase
CC complex during spermatogenesis, an endoprotease complex that
CC catalyzes the intramembrane cleavage of integral membrane proteins
CC such as Notch receptors (lin-12 or glp-1). Involved in spermatid
CC formation during meiosis II. May be required for proper
CC localization of macromolecules that are subject to asymmetric
CC partitioning during spermatogenesis.
CC -!- SUBUNIT: Homodimer. Potential component of the gamma-secretase
CC complex, a complex probably composed of the presenilin homodimer
CC (sel-12, hop-1 or spe-4), nicastrin (aph-2), aph-1 and pen-2
CC (Probable).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Predominantly
CC located in the endoplasmic reticulum and in the cis-Golgi.
CC -!- DEVELOPMENTAL STAGE: Expressed during L4 stage, during
CC spermatogenesis, when hermaphrodites produces sperm.
CC -!- SIMILARITY: Belongs to the presenilin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z14066; CAA78449.1; -.
CC EMBL; Z14067; CAA78450.1; -.
CC EMBL; Z73912; CAA98145.1; -.
CC FIR; T27885; T27885.
CC MEROPS; A22.UEF; -.
CC WormPep; ZK524.1; CE06618.
CC InterPro; IPR006639; Peptidase A22.
CC InterPro; IPR001108; Peptidase_A22A.
CC Pfam; PF01080; Presenilin; 1.
CC SMART; SM00730; PSN; 1.

KW Endoplasmic reticulum; Golgi stack; Sperm; Transmembrane.
 FT TRANSMEM 19 39 POTENTIAL.
 FT TRANSMEM 72 92 POTENTIAL.
 FT TRANSMEM 97 117 POTENTIAL.
 FT TRANSMEM 137 157 POTENTIAL.
 FT TRANSMEM 161 181 POTENTIAL.
 FT TRANSMEM 191 211 POTENTIAL.
 FT TRANSMEM 390 410 POTENTIAL.
 FT TRANSMEM 412 432 POTENTIAL.
 FT TRANSMEM 440 460 POTENTIAL.
 FT MUTAGEN 177 177
 FT MUTAGEN 440 440
 FT MUTAGEN 440 440
 SQ SEQUENCE 465 AA; 51829 MW; 65BE2A4DFD3C844 CRC64;
 Query Match 37.5%; Score 6; DB 1; Length 465;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 NSLPNN 10
 Db 366 NSLPNN 371
 RESULT 7
 ID GLMN MOUSE STANDARD; PRT; 596 AA.
 AC Q8B2ML; Q95L88;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Glomulin (FKBP-associated protein) (FK506-binding protein-associated
 DE protein) (FAP).
 GN GLMN OR FAP48.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CD-1; TISSUE=Embryo;
 RA McIntyre B.A.S., Brouillard P., Aerts V., Gutierrez-Roelens I.,
 RA Vikkula M.
 RA "Glomulin is expressed in vascular smooth muscle cells in the
 RT embryonic and adult mouse.";
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Diencephalon;
 RX MEDLINE=22354683; PubMed=12466851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nixaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusci V., Chochia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gasterland T., Gariboldi M., Gibsi C., Godzik A., Gough J.,
 RA Grամmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RL 60,770 full-length cDNAs.";
 RN Nature 420:563-573 (2002).
 [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2238257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Schenck C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RL human and mouse cDNA sequences.";
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -!- FUNCTION: Essential for normal development of the vasculature. May
 CC represent a naturally occurring ligand of the immunophilins FKBP59
 CC and FKBP12. May function as a membrane anchoring protein. May
 CC stimulate the p70S6K pathway. May inhibit cell proliferation and
 CC increase IL2 production (By similarity).
 CC -!- SUBUNIT: Monomer. Interacts with nonphosphorylated MBT and is
 CC released upon receptor phosphorylation. Interacts with FKBP59 and
 CC FKBP12. Part of a SCF-like complex consisting of CUL7, RBX1, SKP1,
 CC FBXW8 and GLMN (By similarity).
 CC -!- PTM: Phosphorylated on tyrosine residues (By similarity).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ566083; CAD92739.1; -;
 CC EMBL; AK034160; BAC28612.1; -;
 CC EMBL; BC003446; AA03446.1; -;
 CC MGB; MGI:2141180; Fap48.
 CC Coiled coil; Phosphorylation.
 CC DOMAIN 122 125 POLY-LEU.
 CC COILED COIL (POTENTIAL).
 CC FT DOMAIN 193 209 POLY-GLU.
 CC FT DOMAIN 272 277 POLY-GLU.
 CC FT DOMAIN 515 532 COILED COIL (POTENTIAL).
 CC FT CONFLICT 270 270 Y -> S (IN REF. 3).
 CC SQ SEQUENCE 596 AA; 67755 MW; 4C306B16F3C206DB CRC64;
 Query Match 37.5%; Score 6; DB 1; Length 596;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 LPNKA 12
 Db 135 LPNKA 140
 RESULT 8
 YBP3 YEAST STANDARD; PRT; 689 AA.
 ID YBP3_YEAST

AC P38227;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 77.3 kDa protein in FAT1-TCM62 intergenic region.
 GN YBR043C OR YBR0413.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,
 RA Visers S.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
 CC AS THE DRUG RESISTANCE TRANSLUCASE FAMILY). BELONGS TO THE
 CC CARL/CYHR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z35912; CAA84985.1; -;
 DR PIR; S45901; S45901.
 DR GerMOnline; 138586; -;
 DR SGD; S0000247; YBR043C.
 DR InterPro; IPR007114; MFS.
 DR PROSITE; PS50850; MFS; 1.
 DR Hypothetical protein; Transport; Transmembrane.
 FT TRANSMEM 109 131 POTENTIAL.
 FT TRANSMEM 140 163 POTENTIAL.
 FT TRANSMEM 176 193 POTENTIAL.
 FT TRANSMEM 236 256 POTENTIAL.
 FT TRANSMEM 268 283 POTENTIAL.
 FT TRANSMEM 476 493 POTENTIAL.
 FT TRANSMEM 511 532 POTENTIAL.
 FT TRANSMEM 559 577 POTENTIAL.
 FT TRANSMEM 587 609 POTENTIAL.
 FT TRANSMEM 625 642 POTENTIAL.
 FT TRANSMEM 649 668 POTENTIAL.
 FT TRANSMEM 689 AA; 77300 MW; 51420CBB8DB1925F CRC64;
 SQ SEQUENCE 689 AA; 77300 MW; 51420CBB8DB1925F CRC64;
 Query Match 37.5%; Score 6; DB 1; Length 689;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 NSLPNN 10
 Db 20 NSLPNN 25
 RESULT 9
 PTPE MOUSE
 ID -PTPE_MOUSE STANDARD; PRT; 699 AA.
 AC P49446; Q62134; Q62444; Q64496;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Protein-tyrosine phosphatase epsilon precursor (EC 3.1.3.48) (R-PTP-
 DE epsilon).
 GN PTPE OR PTP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N;
 RX MEDLINE=96064677; PubMed=7592814;
 RA Elson A., Leder P.;
 RT "Protein-tyrosine phosphatase epsilon. An isoform specifically
 RT expressed in mouse mammary tumors initiated by v-Ha-ras OR neu.";
 RL J. Biol. Chem. 270:26116-26122 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DBA/2;
 RA Mikoyama Y.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain, and Lung;
 RA Hou E.W., Li S.L.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 224-332 FROM N.A.
 RC STRAIN=BAUB/c; TISSUE=Brain;
 RX MEDLINE=93086603; PubMed=1454056;
 RA Schepens J., Zeeuwen P., Wieringa B., Hendriks W.;
 RT "Identification and typing of members of the protein-tyrosine
 RT phosphatase gene family expressed in mouse brain.";
 RL Mol. Biol. Rep. 16:241-248 (1992).
 RN [5]
 RP SEQUENCE OF 224-332 FROM N.A.
 RC STRAIN=BAUB/c; TISSUE=Brain;
 RX MEDLINE=95134232; PubMed=7832766;
 RA Hendriks W., Schepens J., Brugman C., Zeeuwen P., Wieringa B.;
 RT "A novel receptor-type protein tyrosine phosphatase with a single
 RT catalytic domain is specifically expressed in mouse brain.";
 RL Biochem. J. 305:499-504 (1995).
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
 CC tyrosine + phosphate.
 CC -1- SUBCELLULAR LOCATION: Type 1 membrane protein.
 CC -1- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U35368; AAC52281.1; -;
 DR EMBL; D83484; BAA11927.1; -;
 DR EMBL; U62387; AAB04553.1; -;
 DR EMBL; Z23052; CAA80587.1; -;
 DR EMBL; Z23053; CAA80588.1; -;
 DR PIR; B61180; B61180.
 DR HSP; P18052; 1YFO.
 DR MGD; MGI:97813; Ptpre.
 DR InterPro; IPR000387; TYR phosphatase.
 DR InterPro; IPR000242; Tyr_PP
 DR Pfam; PF00102; Y_phosphatase; 2.
 DR PRINTS; PR00700; PRTYPHPTASE.
 DR SMART; SM00194; PTPC; 2.
 DR PROSITE; PS00383; TYR PHOSPHATASE 1; 2.
 DR PROSITE; PS00566; TYR PHOSPHATASE 2; 2.
 DR PROSITE; PS50055; TYR PHOSPHATASE_PTP; 2.
 KW Glycoprotein; Transmembrane; Hydroxylase; Phosphorylation; Repeat;
 KW Signal.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 699 PROTEIN-TYROSINE PHOSPHATASE EPSILON.
 FT DOMAIN 20 45 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 46 68 POTENTIAL.
 FT DOMAIN 69 699 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 153 392 PROTEIN-TYROSINE PHOSPHATASE 1.
 FT DOMAIN 393 699 PROTEIN-TYROSINE PHOSPHATASE 2.
 FT ACT_SITE 334 334 PHOSPHOCYSTEINE INTERMEDIATE (BY
 FT SIMILARITY).

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE PUF4 protein.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=II125-2B;
 RX MEDLINE=91353086; PubMed=1909073;
 RA Chen W., Balzi E., Capieaux E., Goffeau A.;
 RT "The YGL023 gene encodes a putative regulatory protein.";
 RL Yeast 7:309-312 (1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=II125-2B;
 RX MEDLINE=91353083; PubMed=1882553;
 RA Chen W., Balzi E., Capieaux E., Choder M., Goffeau A.;
 RT "The DNA sequencing of the 17 kb HindIII fragment spanning the LEU1
 and ATE1 loci on chromosome VII from Saccharomyces cerevisiae reveals
 the PDR6 gene, a new member of the genetic network controlling
 pleiotropic drug resistance.";
 RL Yeast 7:287-299 (1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Hebling U., Hofmann B., Delius H.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Is not essential for haploid growth, but may affect
 diploid formation.
 CC -1- SIMILARITY: Contains 8 pumilio repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S57889; AB19616.1; -;
 DR EMBL; S58126; AAD13898.1; -;
 DR EMBL; Z72536; CA95714.1; -;
 DR PIR; S64016; S64016.
 DR Germline; 141062; -;
 DR SGD; S0002982; PUF4.
 DR GO; GO:0007569; P:cell aging; IGI.
 DR GO; GO:0000288; P:mRNA catabolism, deadenylation-dependent; IGI.
 DR GO; GO:0006605; P:protein targeting; IMP.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR001313; Pumilio/Puf.
 DR Pfam; PF00806; PUF; 8.
 DR SMART; SM00025; Pumilio; 8.
 DR Transcription regulation; Zinc-finger; Metal-binding; Repeat.
 FT DOMAIN 8 38
 FT ASP/GLU-RICH (ACIDIC).
 FT POLY-ASN (POTENTIAL ACTIVATING DOMAIN).
 FT C4-TYPE.
 FT PUMILIO 1.
 FT PUMILIO 2.
 FT PUMILIO 3.
 FT PUMILIO 4.
 FT PUMILIO 5.
 FT PUMILIO 6.
 FT PUMILIO 7.
 FT PUMILIO 8.
 FT PUMILIO 9.
 FT A -> R (IN REF. 1 AND 2).
 FT CONFLICT 595 595
 FT SEQUENCE 888 AA; 97797 MW; 659BA1062439F642 CRC64;
 Query Match 37.5%; Score 6; DB 1; Length 888;
 Best Local Similarity 100.0%; Pred. No. 40;

QY 9 NKKASS 14
 Db 195 NKKASS 200
 RESULT 12
 ITAD HUMAN STANDARD; PRT; 1162 AA.
 AC Q13349; Q15575; Q15576;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Integrin alpha-D precursor (Leukointegrin alpha D) (CD11d) (ADB2).
 GN ITGAD.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Splesen;
 RX MEDLINE=96111956; PubMed=8777714;
 RA Van der Vieren M., Le Trong H., Wood C.L., Moore P.F., St John T.,
 RA Staunton D.E., Gallatin W.M.;
 RT "A novel leukointegrin, alpha d beta 2, binds preferentially to ICAM-
 3.";
 RL Immunity 3:683-690 (1995).
 RN [2]
 RP SEQUENCE OF 1-235 FROM N.A.
 RX MEDLINE=20187620; PubMed=10722744;
 RA Noti J.D., Johnson A.K., Dillon J.D.;
 RT "Structural and functional characterization of the leukocyte integrin
 gene CD11d. Essential role of Sp1 and Sp3.";
 RL J. Biol. Chem. 275:8959-8969 (2000).
 RN [3]
 RP SEQUENCE OF 467-537; 571-602; 633-666; 788-834 AND 910-1125 FROM N.A.
 RX MEDLINE=96257236; PubMed=8666289;
 RA Wong D.A., Davis E.M., LeBeau M., Springer T.A.;
 RT "Cloning and chromosomal localization of a novel gene-encoding a human
 beta 2-integrin alpha subunit.";
 RL Gene 171:291-294 (1996).
 RN [4]
 RP INTERACTION WITH VCAM1.
 RX MEDLINE=99059842; PubMed=9841932;
 RA Grayson M.H., Van der Vieren M., Sterbinsky S.A., Michael Gallatin W.,
 RA Hoffman P.A., Staunton D.E., Bochner B.S.;
 RT "alpha2beta2 integrin is expressed on human eosinophils and functions
 as an alternative ligand for vascular cell adhesion molecule 1
 (VCAM-1).";
 RL J. Exp. Med. 188:2187-2191 (1998).
 RN [5]
 RP INTERACTION WITH VCAM1.
 RX MEDLINE=99370002; PubMed=10438935;
 RA Van der Vieren M., Crowe D.T., Hoekstra D., Vazeux R., Hoffman P.A.,
 RA Grayson M.H., Bochner B.S., Gallatin W.M., Staunton D.E.;
 RT "The leukocyte integrin alpha D beta 2 binds VCAM-1: evidence for a
 binding interface between I domain and VCAM-1.";
 RL J. Immunol. 163:1984-1990 (1999).
 CC -1- FUNCTION: INTEGRIN ALPHA-D/BETA-2 IS A RECEPTOR FOR ICAM3 AND
 CLEARING LIPOPROTEINS FROM PLAQUES AND IN PHAGOCYTOSIS OF BLOOD-
 BORNE PATHOGENS, PARTICULATE MATTER, AND SENESCENT ERYTHROCYTES
 FROM THE BLOOD.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-D
 ASSOCIATES WITH BETA-2.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: EXPRESSED MODERATELY ON MYELOMONOCYTIC CELL
 LINES AND SUBSETS OF PERIPHERAL BLOOD LEUKOCYTES AND STRONGLY ON
 TISSUE-SPECIALIZED CELLS, INCLUDING MACROPHAGES FOAM CELLS WITHIN
 ATHEROSCLEROTIC PLAQUES, AND ON SPLENIC RED PULP MACROPHAGES.
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS

CC WITH 1-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
 CC -1- SIMILARITY: Contains 1 VWFA domain.
 CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U37028; AAB38547.1; -
 CC EMBL; U40274; AAB60634.1; -
 CC EMBL; U40275; AAB60635.1; -
 CC EMBL; U40276; AAB60636.1; -
 CC EMBL; U40277; AAB60637.1; -
 CC EMBL; U40278; AAB60638.1; -
 CC EMBL; U40279; AAB60638.1; JOINED.
 CC EMBL; AF187881; AAB60638.1; -
 CC HSSP; F11215; IABX.
 CC Genew; HGNC:6146; ITGAD.
 CC MIM; 602453; -
 CC GO; GO:0008305; C:integrin complex; TAS.
 CC GO; GO:0004895; P:cell adhesion receptor activity; TAS.
 CC GO; GO:0016337; P:cell-cell adhesion; NAS.
 CC GO; GO:0007160; P:cell-matrix adhesion; NAS.
 CC GO; GO:0006955; P:immune response; NAS.
 CC InterPro; IPR000413; Integrin_alpha.
 CC InterPro; IPR002035; VWF_A.
 CC Pfam; PF01839; FG-GAP; 3.
 CC Pfam; PF00357; integrin_A; 1.
 CC Pfam; PF00092; vwa; 1.
 CC PRINTS; PR01185; INTEGRINA.
 CC SMART; PR00453; VWFA DOMAIN.
 CC SMART; SM00191; Int alpha; 4.
 CC SMART; SM00327; VWA; 1.
 CC PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 CC PROSITE; PS50234; VWFA; 1.
 CC Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 CC Signal; Repeat; Calcium;
 CC Magnesium.
 CC SIGNAL 1 17 POTENTIAL.
 CC CHAIN 18 1162 INTEGRIN ALPHA-D.
 CC DOMAIN 18 1100 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 1101 1124 POTENTIAL.
 CC DOMAIN 1125 1162 CYTOPLASMIC (POTENTIAL).
 CC REPEAT 32 85 FG-GAP 1.
 CC REPEAT 7 7 FG-GAP 2.
 CC DOMAIN 150 332 VWFA.
 CC REPEAT 350 400 FG-GAP 3.
 CC REPEAT 401 452 FG-GAP 4.
 CC REPEAT 454 516 FG-GAP 5.
 CC REPEAT 518 576 FG-GAP 6.
 CC REPEAT 581 633 FG-GAP 7.
 CC CA_BIND 465 473 POTENTIAL.
 CC CA_BIND 530 538 POTENTIAL.
 CC CA_BIND 593 601 POTENTIAL.
 CC SITE 1127 1131 GFPR MOTIF.
 CC DISULFID 67 74 BY SIMILARITY.
 CC DISULFID 106 124 BY SIMILARITY.
 CC DISULFID 655 710 BY SIMILARITY.
 CC DISULFID 769 775 BY SIMILARITY.
 CC DISULFID 846 861 BY SIMILARITY.
 CC DISULFID 994 1018 BY SIMILARITY.
 CC DISULFID 1023 1028 BY SIMILARITY.
 CC CARBOHYD 59 59 N-LINKED (GLCNAC...)
 CC CARBOHYD 87 87 N-LINKED (GLCNAC...)
 CC CARBOHYD 99 99 N-LINKED (GLCNAC...)
 CC CARBOHYD 391 391 N-LINKED (GLCNAC...)
 CC CARBOHYD 691 691 N-LINKED (GLCNAC...)
 CC CARBOHYD 733 733 N-LINKED (GLCNAC...)

FT CARBOHYD 873 873 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 957 957 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1046 1046 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 500 500 MISSING (IN REF. 2).
 FT CONFLICT 515 518 GHPW -> ATP (IN REF. 2).
 FT CONFLICT 825 825 L -> V (IN REF. 2).
 FT CONFLICT 984 984 V -> A (IN REF. 2).
 SQ SEQUENCE 1162 AA; 126885 MW; F296A1A35455D77D CRC64;
 Query Match 37.5%; Score 6; DB 1; Length 1162;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 9 NKKASS 14
 Db 902 NKKASS 907
 ID GBG7 MOUSE STANDARD; PRT; 42 AA.
 AC Q61016;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Guanine nucleotide-binding protein G(I)/G(S)/G(O) gamma-7 subunit
 DE (Fragment).
 GN GNG7 OR GNGT7
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CF-1 / Harlan;
 RX MEDLINE=97011591; PubMed=8858601;
 RA Williams C.J., Schultz R.M., Kopf G.S.;
 RT "G protein gene expression during mouse oocyte growth and maturation,
 RT and preimplantation embryo development.";
 RL Mol. Reprod. Dev. 44:315-323(1996).
 CC -1- FUNCTION: Guanine nucleotide-binding proteins (G proteins) are
 CC involved as a modulator or transducer in various transmembrane
 CC signaling systems. The beta and gamma chains are required for the
 CC GTPase activity, for replacement of GDP by GTP, and for G protein-
 CC effector interaction.
 CC -1- SUBUNIT: G proteins are composed of 3 units, alpha, beta and
 CC gamma.
 CC -1- SIMILARITY: Belongs to the G protein gamma family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U38499; AAB01730.1; -
 CC MGD; MGI:95787; Gng7.
 CC InterPro; IPR001770; G-gamma.
 CC Pfam; PF00631; G-gamma; 1.
 CC PRINTS; PR00321; GPROTEING.
 CC ProDom; PD003783; G-gamma; 1.
 CC PROSITE; PS50058; G-PROTEIN GAMMA; 1.
 KW Transducer; Prenylation; Lipoprotein; Multigene family.
 FT NON_TER 1 1
 FT NON_TER 42 42
 FT SEQUENCE 42 AA; 4695 MW; C85E514E69D95F4 CRC64;
 Query Match 31.2%; Score 5; DB 1; Length 42;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Brain;
RX  MEDLINE=96140645; PubMed=8548815;
RA  Koelle M.R., Horvitz H.R.;
RT  "EGT-10 regulates G protein signaling in the C. elegans nervous
RT  system and shares a conserved domain with many mammalian proteins.";
RL  Cell 84:115-125(1996).
CC  -1- FUNCTION: INHIBITS SIGNAL TRANSDUCTION BY INCREASING THE GTPASE
CC  ACTIVITY OF G PROTEIN ALPHA SUBUNITS THEREBY DRIVING THEM INTO
CC  THEIR INACTIVE GDP-BOUND FORM. ASSOCIATES SPECIFICALLY WITH THE
CC  ACTIVATED FORMS OF THE G PROTEIN SUBUNITS G(1)-ALPHA AND G(2)-
CC  ALPHA BUT FAILS TO INTERACT WITH THE STRUCTURALLY AND FUNCTIONALLY
CC  DISTINCT G(5)-ALPHA SUBUNIT. ACTIVITY ON G(2)-ALPHA IS INHIBITED
CC  BY PALMITOYLATION OF THE G-PROTEIN (BY SIMILARITY).
CC  -1- SIMILARITY: Contains 1 RGS domain.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  ENBL; U32437; AAC52374.1; -
DR  InterPro; IPR000342; Regl_Gproteins.
DR  Pfam; PF00615; RGS; 1.
DR  PRINTS; PR01301; RGS-PROTEIN.
DR  ProDom; PD001580; Regl_Gproteins; 1.
DR  PROSITE; PS50132; RGS; 1.
DR  XW  Signal transduction inhibitor; Lipoprotein; Palmitate.
FT  NON_TER 1 1
FT  DOMAIN <1 >66 RGS.
FT  LIPID 3 3 S-palmitoyl cysteine (By similarity).
FT  NON_TER 66 66
FT  SQ SEQUENCE 66 AA; 7776 MW; 052DBC45DA024251 CRC64;
Query Match 31.2%; Score 5; DB 1; Length 66;
Best local Similarity 100.0%; Pred. No. 45;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 NKASS 14
DB 29 NKASS 33

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:47:36 ; Search time 33.3333 Seconds

(without alignments)
151.449 Million cell updates/sec

Title: US-09-528-682-4_COPY_64_79

Perfect score: 16

Sequence: 1 VREFNSLPNNKASDPT 16

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_page:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	43.8	37	5 Q9NFF7	Q9nff7 plasmodium
2	7	43.8	90	5 Q8IK24	Q8ik24 plasmodium
3	7	43.8	105	17 Q9YBU3	Q9ybu3 aeropyrum p
4	6	37.5	97	16 Q8KES2	Q8kes2 chlorobium
5	6	37.5	100	8 Q8LVU5	Q8lvu5 chaetophae
6	6	37.5	124	3 Q07438	Q07438 saccharomyc
7	6	37.5	130	3 Q7Z994	Q7z994 schizosach
8	6	37.5	141	9 Q9AZF1	Q9azf1 bacterioph
9	6	37.5	141	16 Q9C155	Q9c155 lactococcus
10	6	37.5	155	4 Q96P81	Q96p81 homo sapien
11	6	37.5	162	13 Q7SYW7	Q7syw7 xenopus lae
12	6	37.5	191	16 Q97IT3	Q97it3 clostridium
13	6	37.5	224	11 Q8R3Y3	Q8r3y3 mus musculu
14	6	37.5	257	16 Q7WGF3	Q7wgf3 bordetella
15	6	37.5	257	16 Q7W4X5	Q7w4x5 bordetella
16	6	37.5	257	16 Q7VUI6	Q7vui6 bordetella

17	6	37.5	265	2 Q8KPZ6	Q8kpz6 haemophilus
18	6	37.5	282	5 Q8II65	Q8ii65 plasmodium
19	6	37.5	297	11 Q63476	Q63476 rattus norv
20	6	37.5	341	5 Q25266	Q25266 leishmania
21	6	37.5	341	5 Q9GNM8	Q9gnm8 leishmania
22	6	37.5	345	10 Q84PP2	Q84pp2 glycine max
23	6	37.5	360	16 Q7V9C0	Q7v9c0 prochloroco
24	6	37.5	369	13 Q7ZXD5	Q7zxd5 xenopus lae
25	6	37.5	380	10 Q8SD5	Q8sds5 guillardia
26	6	37.5	399	16 Q8YK50	Q8yk50 anabaena sp
27	6	37.5	407	16 Q7UI35	Q7ui35 rhodospirell
28	6	37.5	469	12 Q8JNY0	Q8jny0 influenza a
29	6	37.5	469	12 Q8QLV9	Q8qlv9 influenza a
30	6	37.5	469	12 Q8JUH0	Q8jhu0 influenza a
31	6	37.5	469	12 Q8JVM2	Q8jvm2 influenza a
32	6	37.5	469	12 Q8JNY2	Q8jny2 influenza a
33	6	37.5	469	12 Q8VW8	Q8vuw8 influenza a
34	6	37.5	469	12 Q8VVV9	Q8vvv9 influenza a
35	6	37.5	469	12 Q8JNY1	Q8jny1 influenza a
36	6	37.5	469	12 Q8JNY3	Q8jny3 influenza a
37	6	37.5	469	12 Q8JV11	Q8jv11 influenza a
38	6	37.5	469	12 Q8QLP8	Q8qlp8 influenza a
39	6	37.5	469	12 Q8AZB4	Q8azb4 influenza a
40	6	37.5	469	12 Q8AZB3	Q8azb3 influenza a
41	6	37.5	472	16 Q8UEB2	Q8ueb2 agrobacteri
42	6	37.5	502	5 Q16676	Q16676 caenorhabdi
43	6	37.5	510	2 Q8GQC2	Q8gqc2 lactobacill
44	6	37.5	515	11 Q8K082	Q8k082 mus musculu
45	6	37.5	521	11 Q7TMR2	Q7tmr2 mus musculu

ALIGNMENTS

RESULT 1

Q9NFF7 ID Q9NFF7 PRELIMINARY; PRT; 37 AA.
AC Q9NFF7; 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 25, Last annotation update)
DE Putative acyl-CoA-binding protein (Fragment).
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=21036610; PubMed=11163452;
RA Spielmann T., Beck H.P.;
RT "Analysis of stage-specific transcription in Plasmodium falciparum
RT reveals a set of genes exclusively transcribed in ring stage
parasites.";
RL Mol. Biochem. Parasitol. 111:453-458(2000).
DR EMBL; AJ280941; CAB92949.1; -;
DR GO; GO:0000062; F:acyl-CoA binding; IEA.
DR InterPro; IPR000582; Ac_coa_bind_prot.
FT NON_TER 1
FT NON_TER 37
SQ SEQUENCE 37 AA; 4369 MW; 0877BC25708865BA CRC64;

Query Match 43.8%; Score 7; DB 5; Length 37;
Best Local Similarity 100.0%; Pred.No. 1.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NSLPNNK 11

|||||
14 NSLPNNK 20

RESULT 2

ID Q8IK24 PRELIMINARY; PRT; 90 AA.
ID Q8IK24

Q81K24;
 01-WAR-2003 (TrEMBLrel. 23, Created)
 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)
 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 Acyl CoA binding protein, putative.
 GN PF10_0016.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 SEQUENCE FROM N.A.

AC STRAIN=307;
 RC MEDLINE=22255705; PubMed=12368864;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
 RA Fertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
 RA McFadden G.I., Cummings I.M., Subramanian G.M., Mungall C.,
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrell B.;
 RT "Genome sequence of the human malaria parasite Plasmodium
 RT falciparum.";
 RL Nature 419:498-511(2002).
 DR EMBL; AS014829; AAN35214.1; -;
 DR GO; GO:0000622; F:acyl-CoA binding; IEA.
 DR InterPro; IPR000582; Ac_coA_bind_prot.
 DR Pfam; PF00887; ACBP; 1.
 DR PRINTS; PR00689; AC0ABINDINGP.
 DR SQ SEQUENCE 90 AA; 10767 MW; 3A6778530FFBFA16 CRC64;

Query Match 43.8%; Score 7; DB 5; Length 90;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NSLPNKK 11
 |||||
 DB 13 NSLPNKK 19

RESULT 3
 Q9YBU3 PRELIMINARY; PRT; 105 AA.
 ID Q9YBU3
 AC Q9YBU3;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein APE1506.
 GN APE1506.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
 OC Desulfurococaceae; Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 SEQUENCE FROM N.A.

AC STRAIN=KJ;
 RC MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaï A., Kosugi H.,
 RA Hoshizawa A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 DR EMBL; AP000061; BAA0505.1; -;
 DR PIR; C72631; C72631.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 105 AA; 11568 MW; 5B33500281A43CC0 CRC64;

Query Match 43.8%; Score 7; DB 17; Length 105;

Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 LPNNKAS 13
 |||||
 DB 57 LPNNKAS 63

RESULT 4
 Q8KES2 PRELIMINARY; PRT; 97 AA.
 ID Q8KES2
 AC Q8KES2;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Cold shock-like protein CapG.
 GN CPBG OR CT0610.
 OS Chlorobium tepidum.
 OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
 OC Chlorobium.
 OX NCBI_TaxID=1097;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=TLS / ATCC 49652 / DSM 12025;
 RX MEDLINE=22103685; PubMed=12093901;
 RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
 RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
 RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
 RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
 RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
 RA Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
 RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
 RA "The complete genome sequence of Chlorobium tepidum TLS, a
 RT photosynthetic, anaerobic, green-sulfur bacterium.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
 DR EMBL; AS012834; AAM71852.1; -;
 DR TIGR; CT0610; -;
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR002059; Cold shock.
 DR InterPro; IPR008994; Nucleic_acid_OB.
 DR Pfam; PF00313; CSP; 1.
 DR PRINTS; PR00050; COLDSHOCK.
 DR PRODOM; PD000621; Cold_shock; 1.
 DR SMART; SM00357; CSP; 1.
 DR PROSITE; PS00352; COLD_SHOCK; 1.
 KW Complete proteome.
 SQ SEQUENCE 97 AA; 10578 MW; 56D77AE0F59C1BE3 CRC64;

Query Match 37.5%; Score 6; DB 16; Length 97;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8
 |||||
 DB 91 EFNSLP 96

RESULT 5
 Q8LVU5 PRELIMINARY; PRT; 100 AA.
 ID Q8LVU5
 AC Q8LVU5;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein.
 GN ORF100.
 OS Chaetosphaeridium globosum.
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Coleochaetales;
 OC Chaetosphaeridiaceae; Chaetosphaeridium.
 OX NCBI_TaxID=96477;
 RN [1]

```

RP SEQUENCE FROM N.A.
RX MEDLINE=22177139; PubMed=12161560;
RA Turmel M., Otis C., Lemieux C.;
RT "The chloroplast and mitochondrial genome sequences of the charophyte
RT Chaetopharidium globosum: Insights into the timing of the events
RT that restructured organelle DNAs within the green algal lineage that
RT led to land plants.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:11275-11280(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Lemieux C., Otis C., Turmel M.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF494278; AAM96592.1; -
DR EMBL; AF494278; AAM96594.1; -
DR GO; GO:0009507; C:chloroplast; IEA.
KW Hypothesis; C:chloroplast.
SQ SEQUENCE 100 AA; 11731 MW; 22686233199CC9CB2 CRC64;

Query Match 37.5%; Score 6; DB 3; Length 100;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 LPNKA 12
Db 40 LPNKA 45

RESULT 6
Q07438 PRELIMINARY; PRT; 124 AA.
AC Q07438;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE ORF YDL071C.
GN YDL071C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Schizosaccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Bloeker H., Brandt P.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z74119; CAA98635.1; -
DR PIR; S67606; S67606.
DR SGD; S0002229; YDL071C.
SQ SEQUENCE 124 AA; 13602 MW; 3FD5532158BDF3F3 CRC64;

Query Match 37.5%; Score 6; DB 3; Length 124;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 SLPNK 11
Db 40 SLPNK 45

RESULT 7
Q72994 PRELIMINARY; PRT; 130 AA.
AC Q72994;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Spn7 protein (SPBC21.08c protein) (SPBC19F8.01c protein)
DE (fragment).
GN SPBC19F8.01C OR SPBC21.08C OR SPN7.
OS Schizosaccharomycos pombe (Fission yeast).

```

```

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972h-;
MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Brooks J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Collins M., Brown D., Brown S., Chillingworth T., Churcher C.,
RA Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Hollroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vansteels E., Rieger M., Schafer M., Muller-Auer S.,
RA Gabel C., Fuchs M., Dusterhoft A., Fritzc C., Holzer E., Moestl D.,
RA Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R.,
RA Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Usery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomycos pombe.";
RL Nature 415:871-880(2002).
DR EMBL; AL035537; CAD99124.1; -
FT NON_TER 1
SQ SEQUENCE 130 AA; 14891 MW; 80DAD75B551AB02A CRC64;

Query Match 37.5%; Score 6; DB 3; Length 130;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 NSLPNN 10
Db 100 NSLPNN 105

RESULT 8
Q9AZF1 PRELIMINARY; PRT; 141 AA.
AC Q9AZF1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Orf9.
GN ORF9.
OS Bacteriophage b1312.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=151539;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21113149; PubMed=11160885;
RA Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., Chopin M.-C.;
RT "Analysis of six prophages in Lactococcus lactis I11403: different
RT genetic structure of temperate and virulent phage populations.";
RL Nucleic Acids Res. 29:644-651(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., Chopin M.-C.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF232673; AAK08462.1; -
SQ SEQUENCE 141 AA; 16760 MW; 04DAE642467D7676 CRC64;

```

Query Match 37.5%; Score 6; DB 9; Length 141;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 REFNSL 7
 Db 114 REFNSL 119

RESULT 9

Q9C155 PRELIMINARY; PRT; 141 AA.
 AC Q9C155;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Prophage pS2 protein 09.
 GN PS209 OR L0511.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 OX NCBI_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IL1403;
 RX MEDLINE=21235186; PubMed=11337471;
 RA Bolotin A., Wincker P., Mauger S., Jallion O., Malarne K.,
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 RT lactis ssp. lactis IL1403";
 RL Genome Res. 11:731-753(2001).
 DR EMBL; AE006286; AAK04609.1; -.
 DR PIR; G86688; G86688.
 KW Complete proteome.
 SQ SEQUENCE 141 AA; 16760 MW; 04DAB642467D7676 CRC64;

Query Match 37.5%; Score 6; DB 16; Length 141;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 REFNSL 7
 Db 114 REFNSL 119

RESULT 10

Q96P81 PRELIMINARY; PRT; 155 AA.
 AC Q96P81;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PTPRE (Fragment).
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Zhao Y., Sun D., Dai S.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF406557; AAL01375.1; -.
 FT NON TER 155 155
 SQ SEQUENCE 155 AA; 8A7DB29677F06E00 CRC64;

Query Match 37.5%; Score 6; DB 4; Length 155;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8
 Db 138 EFNSLP 143

RESULT 11

Q7SYW7 PRELIMINARY; PRT; 162 AA.
 AC Q7SYW7;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Whole;
 RX MEDLINE=22341132; PubMed=12454917;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT initiative";
 RL Dev. Dyn. 225:384-391(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Whole;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Sapletchenko M., Soares M.B., Bonaldo M.F., Casavant T.B., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Frange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Rosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Whole;
 RA Klein S., Strausberg R.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC054238; AAH54238.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 162 AA; 17827 MW; D7A1DAD6E4FAP8B CRC64;

Query Match 37.5%; Score 6; DB 13; Length 162;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 SLPNNK 11
 Db 28 SLPNNK 33

RESULT 12

Q97IT3 PRELIMINARY; PRT; 191 AA.
 ID Q97IT3
 AC Q97IT3;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Predicted acetyltransferase.
 GN CAC1557
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=1466286;
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatunov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RA "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum.";
 RL J. Bacteriol. 183:4823-4838(2001).
 DR EMBL; AB007665; AAK79524.1; -;
 DR PIR; A97092; A97092.
 DR GO; GO:0008080; F.N-acetyltransferase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR InterPro; IPR000182; GCNSacetyl_trans.
 DR Pfam; PF00583; Acetyltransf; 1.
 KW Transferase; Complete proteome.
 SQ SEQUENCE 191 AA; 22201 MW; 40A0ECE5B914CF80 CRC64;

Query Match 37.5%; Score 6; DB 16; Length 191;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PNNKAS 13
 |||||
 Db 135 PNNKAS 140

RESULT 13
 Q8R3Y3 PRELIMINARY; PRT; 224 AA.
 AC Q8R3Y3;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Similar to unknown (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Strausberg R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC023375; AAK23375.1; -;
 FT NON TER 1
 SQ SEQUENCE 224 AA; 24056 MW; 2C2431F8573AAE93 CRC64;

Query Match 37.5%; Score 6; DB 11; Length 224;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FNSLPN 9
 |||||
 Db 179 FNSLPN 184

RESULT 14
 Q7MGF3 PRELIMINARY; PRT; 257 AA.
 AC Q7MGF3;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative molybdate-binding periplasmic protein.
 GN BB3966.

OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=518;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271;
 RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Harris B., Quail M.A.,
 RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
 RA Rabinowitz E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RA "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640449; CAE34329.1; -;
 KW Complete proteome.
 SQ SEQUENCE 257 AA; 28052 MW; 1CF6D4F840071D28 CRC64;

Query Match 37.5%; Score 6; DB 16; Length 257;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LPNNKA 12
 |||||
 Db 77 LPNNKA 82

RESULT 15
 Q7W4X5 PRELIMINARY; PRT; 257 AA.
 AC Q7W4X5;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative molybdate-binding periplasmic protein.
 GN BPP3531.
 OS Bordetella parapertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=12822 / ATCC BAA-587;
 RX MEDLINE=22827954; PubMed=12910271;
 RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Harris B., Quail M.A.,
 RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
 RA Rabinowitz E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RA "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640433; CAE38815.1; -;
 KW Complete proteome.

Query Match 37.5%; Score 6; DB 16; Length 257;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LPNNKA 12

Db | || || ||
 77 LPNKA 82

Search completed: April 23, 2004, 14:54:11
Job time : 35.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:45:50 ; Search time 48 Seconds
(without alignments)
94.183 Million cell updates/sec

Title: US-09-528-682-4_COPY_64_79
Perfect score: 16
Sequence: 1 VRENSLPNNKASDST 16

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq 29Jan04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	37.5	28	AAR40085	Aar40085 Hib OMP P
2	6	37.5	60	AGG01836	Agg01836 Human sec
3	6	37.5	141	AB53820	Ab53820 Lactococc
4	6	37.5	297	ADD45312	Add45312 Rat Prote
5	6	37.5	297	ADE56322	Ad56322 Rat Prote
6	6	37.5	341	AAR07043	Aar07043 P3 gene p
7	6	37.5	411	AAR34584	Aar34584 Mutant hu
8	6	37.5	465	AAY51378	Aay51378 C. elegan
9	6	37.5	573	AAU08491	Aau08491 Mouse VMG
10	6	37.5	596	AD60772	Ad60772 Rat Prote
11	6	37.5	642	ABU08103	Abu08103 Human kin
12	6	37.5	642	ABU04245	Abu04245 Human exp
13	6	37.5	642	ABU04254	Abu04254 Human exp
14	6	37.5	699	AAW27226	Aaw27226 Mouse pro
15	6	37.5	699	AAW94028	Aaw94028 Mouse pro
16	6	37.5	699	AAU01460	Aau01460 Mouse pro
17	6	37.5	699	ABB57081	Abb57081 Mouse isc
18	6	37.5	700	AAM23827	Aam23827 Human EST
19	6	37.5	700	AAM38831	Aam38831 Human pol
20	6	37.5	700	ABU04246	Abu04246 Human exp
21	6	37.5	700	ABU04250	Abu04250 Human exp
22	6	37.5	700	ABU04253	Abu04253 Human exp
23	6	37.5	700	ABU04249	Abu04249 Human exp
24	6	37.5	700	ABU04252	Abu04252 Human exp
25	6	37.5	717	AAM40617	Aam40617 Human pol

26	6	37.5	717	6	ABU04251	Abu04251 Human exp
27	6	37.5	934	4	ABB69655	Abb69655 Drosophil
28	6	37.5	963	7	ADC31094	Adc31094 Human nov
29	6	37.5	1011	5	ABP73711	Abp73711 Candida a
30	6	37.5	1161	2	AAR78166	Aar78166 Human bet
31	6	37.5	1161	2	AAW23049	Aaw23049 Human bet
32	6	37.5	1161	2	AAW23064	Aaw23064 Human bet
33	6	37.5	1161	2	AAW57491	Aaw57491 Human bet
34	6	37.5	1161	2	AAW65106	Aaw65106 Human bet
35	6	37.5	1161	2	AAW65089	Aaw65089 Human bet
36	6	37.5	1161	2	AAW72837	Aaw72837 Human alp
37	6	37.5	1161	2	AAW72825	Aaw72825 Human alp
38	6	37.5	1161	2	AAW73342	Aaw73342 Human alp
39	6	37.5	1161	2	AAW73343	Aaw73343 Human alp
40	6	37.5	1161	3	AB073359	Ab073359 Human alp
41	6	37.5	1161	3	AB07376	Ab07376 Human alp
42	6	37.5	1161	5	ABG61468	Abg61468 Human Bet
43	6	37.5	1161	5	ABG61485	Abg61485 Human Bet
44	5	31.2	9	6	ABJ38077	Abj38077 Human cyt
45	5	31.2	12	7	ADC27956	Adc27956 Synthetic

ALIGNMENTS

RESULT 1						
AAR40085						
ID	AAR40085 standard; peptide; 28 AA.					
XX						
AC	AAR40085;					
XX						
DT	25-MAR-2003 (revised)					
DT	04-FEB-1994 (first entry)					
XX						
DE	Hib OMP P2 peptide OMP2-6U (148-174).					
XX						
KW	Haemophilus influenzae; type b; Hib; outer membrane protein; P1; P2; P6;					
KW	vaccine; antibody; detection; lipoglycopeptide conjugate; immunogen.					
XX	Synthetic.					
XX						
Key	Location/Qualifiers					
FT	Misc-difference 1					
FT	/note= "May be absent"					
FT	Misc-difference 16..17					
FT	/note= "Residues different from those found in P2 protein of H. influenzae strain 1H"					
FT						
FT	Misc-difference 20					
FT	/note= "Residue different from those found in P2 protein of H. influenzae strain 1H"					
FT						
FT	Misc-difference 23					
FT	/note= "Residue different from those found in P2 protein of H. influenzae strain 1H"					
FT						
XX	W09315205-A2.					
XX						
PD	05-AUG-1993.					
XX						
PF	03-FEB-1993; 93WO-CA000041.					
XX						
XX	03-FEB-1992; 92GB-00002219.					
XX						
PA	(CONN-) CONNAUGHT LAB LTD.					
XX						
PI	Chong P, Kandil A, Sia C, Klein MH;					
XX	WPI; 1993-258681/32.					
XX						
PT	Synthetic Haemophilus influenzae conjugate vaccine - comprising T-helper cell determinants and B-cell epitope(s) linked to synthetic oligo:saccharide(s).					
XX						
PS	Table 2; Page 50; 99pp; English.					

XX The sequences given in AAR40053-101 are peptide fragments derived from
 CC the Haemophilus influenzae type b (Hib) outer membrane proteins P1, P2
 CC and P6. These peptides may be used in a vaccine against Hib infection and
 CC antibodies against these peptides may be used in test kits to detect H.
 CC influenzae in a sample. The vaccine may further comprise an immunogenic or
 CC immunostimulatory molecule or the peptides may be modified with lipids,
 CC or linked to synthetic PRP as synthetic lipoglycopeptide conjugates to
 CC produce alternative vaccines. (Updated on 25-MAR-2003 to correct PN
 CC field.)
 XX

SQ Sequence 28 AA;

Query Match 37.5%; Score 6; DB 2; Length 28;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LPNKA 12
 |||||
 Db 20 LPNKA 25

RESULT 2

AAG01836
 ID AAG01836 standard; protein; 60 AA.

XX AC AAG01836;

XX 06-OCT-2000 (first entry)

DT Human secreted protein, SEQ ID NO: 5917.

DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.

XX Homo sapiens.

XX BP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-00200610.

XX 26-FEB-1999; 99US-0122487P.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX N-PSDB; AAC01842.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX Claim 13; SEQ ID NO 5917; 71pp + Sequence Listing; English.

XX The present sequence is a polypeptide encoded by one of a large number of
 CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
 CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
 CC tissues. EST sequences usually correspond mainly to the 3' untranslated
 CC region (UTR) of the mRNA because they are often obtained from oligo-dr
 CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
 CC sequences derived from the 5' ends of mRNAs and even in those cases where
 CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
 CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
 CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
 CC are also used in diagnostic, forensic, gene therapy and chromosome
 CC mapping procedures. They are used to obtain upstream regulatory sequences
 CC and to design expression and secretion vectors

XX Sequence 60 AA;

Query Match 37.5%; Score 6; DB 3; Length 60;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 SLPNNK 11
 |||||
 Db 16 SLPNNK 21

RESULT 3

ABB53820
 ID ABB53820 standard; protein; 141 AA.

XX AC ABB53820;

XX 29-AUG-2003 (revised)

DT 16-MAY-2002 (first entry)

XX Lactococcus lactis protein ps209.

XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

XX Lactococcus lactis; IL1403.

XX FR2807446-A1.

XX 12-OCT-2001.

XX 11-APR-2000; 2000FR-00004630.

XX 11-APR-2000; 2000FR-00004630.

XX (INRG) INRA INST NAT RECH AGRONOMIQUE.

XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;

XX WPI; 2002-043418/06.

XX New nucleotide sequence useful in the identification of Lactococcus
 PT lactis and related species.

XX Claim 6; SEQ ID NO 522; 2504pp; French.

XX The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic
 CC acid sequence is useful in the detection and/or amplification of nucleic
 CC acid sequence, particularly to identify Lactococcus lactis or related
 CC species. The proteins of the invention are useful for the biosynthesis or
 CC biodegradation of a composition of interest. The invention helps research
 CC in lactic bacteria, particularly useful in the production of yogurt and
 CC cheese. Note: The sequence data for this patent is based on equivalent
 CC patent WO200177334 (published 18-OCT-2001) which is available in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
 CC standardise OS field)

XX Sequence 141 AA;

Query Match 37.5%; Score 6; DB 5; Length 141;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 REFNSL 7
 |||||
 Db 114 REFNSL 119

RESULT 4

ADD45312
 ID ADD45312 standard; protein; 297 AA.

XX AC ADD45312;

XX 29-JAN-2004 (first entry)
 XX Rat Protein BAA11433, SEQ ID NO 10745.
 XX
 XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 XX
 XX Rattus norvegicus.
 XX WO2003016475-A2.
 XX
 XX 27-FEB-2003.
 XX
 XX 14-AUG-2002; 2002WO-US025765.
 XX
 XX 14-AUG-2001; 2001US-0312147P.
 XX 01-NOV-2001; 2001US-0346382P.
 XX 26-NOV-2001; 2001US-0333347P.
 XX
 XX (GEO) GEN HOSPITAL CORP.
 XX (FARB) BAYER AG.
 XX
 XX Woolf C, D'urso D, Befort K, Costigan M;
 XX WPI; 2003-268312/26.
 XX GENBANK; BAA11433.
 XX
 XX New composition comprising two or more isolated polypeptides, useful for
 XX preparing a medicament for treating pain in an animal.
 XX
 XX Claim 1; Page; 1017pp; English.
 XX
 XX The invention discloses a composition comprising two or more isolated rat
 XX or human polynucleotides or a polynucleotide which represents a fragment,
 XX derivative or allelic variation of the nucleic acid sequence. Also
 XX claimed are a vector comprising the novel polynucleotide, a host cell
 XX comprising the vector, a method for identifying a nucleotide sequence
 XX kit to perform the method, an array, a method for identifying an agent
 XX that increases or decreases the expression of the polynucleotide sequence
 XX that is differentially expressed in neuronal tissue of a first animal
 XX subjected to pain, a method for identifying a compound which regulates
 XX the expression of a polynucleotide sequence which is differentially
 XX expressed in an animal subjected to pain, a method for identifying a
 XX compound that regulates the activity of one or more of the
 XX polynucleotides, a method for producing a pharmaceutical composition, a
 XX method for identifying a compound or small molecule that regulates the
 XX activity in an animal of one or more of the polypeptides given in the
 XX specification, a method for identifying a compound useful in treating
 XX pain and a pharmaceutical composition comprising the one or more
 XX polypeptides or their antibodies. The polynucleotide or the compound that
 XX modulates its activity is useful for preparing a medicament for treating
 XX pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 XX therapy). The sequence presented is a rat protein (shown in Table 2 of
 XX the specification) which is differentially expressed during pain. Note:
 XX The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic form directly from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 297 AA;
 XX
 XX Query Match 37.5%; Score 6; DB 7; Length 297;
 XX Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8
 DB 137 EFNSLP 142

RESULT 5

ADE56322
 ID ADE56322 standard; protein; 297 AA.
 XX
 XX ADE56322;
 AC
 XX 29-JAN-2004 (first entry)
 DT
 XX Rat Protein BAA11433, SEQ ID NO 2174.
 DE
 XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 XX
 XX Rattus norvegicus.
 OS
 XX WO2003016475-A2.
 XX
 XX 27-FEB-2003.
 PD
 XX 14-AUG-2002; 2002WO-US025765.
 PF
 XX 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 XX (GEO) GEN HOSPITAL CORP.
 XX (FARB) BAYER AG.
 XX
 XX Woolf C, D'urso D, Befort K, Costigan M;
 PI WPI; 2003-268312/26.
 XX GENBANK; BAA11433.
 DR
 XX New composition comprising two or more isolated polypeptides, useful for
 XX preparing a medicament for treating pain in an animal.
 XX
 XX Claim 1; Page; 1017pp; English.
 PS
 XX The invention discloses a composition comprising two or more isolated rat
 XX or human polynucleotides or a polynucleotide which represents a fragment,
 XX derivative or allelic variation of the nucleic acid sequence. Also
 XX claimed are a vector comprising the novel polynucleotide, a host cell
 XX comprising the vector, a method for identifying a nucleotide sequence
 XX kit to perform the method, an array, a method for identifying an agent
 XX that increases or decreases the expression of the polynucleotide sequence
 XX that is differentially expressed in neuronal tissue of a first animal
 XX subjected to pain, a method for identifying a compound which regulates
 XX the expression of a polynucleotide sequence which is differentially
 XX expressed in an animal subjected to pain, a method for identifying a
 XX compound that regulates the activity of one or more of the
 XX polynucleotides, a method for producing a pharmaceutical composition, a
 XX method for identifying a compound or small molecule that regulates the
 XX activity in an animal of one or more of the polypeptides given in the
 XX specification, a method for identifying a compound useful in treating
 XX pain and a pharmaceutical composition comprising the one or more
 XX polypeptides or their antibodies. The polynucleotide or the compound that
 XX modulates its activity is useful for preparing a medicament for treating
 XX pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 XX therapy). The sequence presented is a rat protein (shown in Table 2 of
 XX the specification) which is differentially expressed during pain. Note:
 XX The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic form directly from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 297 AA;

Query Match 37.5%; Score 6; DB 7; Length 297;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8
 |||||

XX	PD	19-MAY-1993.
XX	PF	06-OCT-1992; 92EP-00117000.
XX	PR	07-OCT-1991; 91JP-00289257.
XX	PA	(GREC) GREEN CROSS CORP.
XX	PI	Tanabe T, Morita M, Hirose M, Amatsuji Y;
XX	DR	WPI; 1993-160551/20.
XX	DR	N-PSDB; AAQ41450.
XX	PT	New human pro-uokinase mutants with thrombolytic activity - have a neutral aminoacid in the epidermal growth factor region replaced with a basic aminoacid, or an acid residue replaced with a non-acidic residue.
XX	PS	Claim 1; Page 17-20; 38pp; English.
XX	CC	The sequence is that of a mutant human prourokinase, in which a neutral amino acid in the epidermal growth region has been replaced with a basic amino acid, or an acidic amino acid has been replaced by a non-acidic amino acid. Preferred replacements are 16Gly->Lys, 38Gly->Lys and 45Asp->Asn. The mutant has an increased half-life in blood as compared to the prior art mutant with a deleted RGF region. It has improved affinity for fibrin, and has other features the same as human prourokinase. (Updated on 25-MAR-2003 to correct PN field.)
XX	SQ	Sequence 411 AA;
	Query Match	37.5%; Score 6; DB 2; Length 411;
	Best Local Similarity	100.0%; Pred. No. 2.9e+02;
	Matches	6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	11 KASDDT 16	
Db	61 KASDDT 66	
	RESULT 8	
AAV51378	ID	AAV51378 standard; protein; 465 AA.
XX	AC	RAY51378;
XX	DT	04-MAY-2000 (first entry)
XX	DE	C. elegans SPE-4 protein.
XX	XW	SPE-4; immunogen; diagnosis; nematode; Alzheimer's disease; brain; muscle; peripheral blood cell; neuroprotectant; nootropic.
XX	OS	Caenorhabditis elegans.
XX	PN	US6019974-A.
XX	PD	01-FEB-2000.
XX	PF	24-JAN-1997; 97US-00788231.
XX	PR	26-JAN-1996; 96US-0010672P.
XX	PA	(UYEM-) UNIV EMORY.
XX	PI	L'hernault SW;
XX	DR	WPI; 2000-146863/13.
XX	DR	N-PSDB; AAZ88555.
XX	PT	Immunogenic composition comprising an antigenic component, useful for diagnosing Alzheimer's disease.
XX	PN	PDE41952-A1

Db	137 EFNSLP 142	
	RESULT 6	
AAR07043	ID	AAR07043 standard; protein; 341 AA.
XX	AC	AAR07043;
XX	DT	25-MAR-2003 (revised)
XX	DT	04-DEC-1990 (first entry)
XX	DE	P3 gene product of Haemophilus strain 8358 (OMP subtype 6U).
XX	DE	P3 gene; strain 1H; influenza; vaccine; ds.
XX	KW	Haemophilus influenzae.
XX	OS	Haemophilus influenzae.
XX	PN	EP378929-A.
XX	PD	25-JUL-1990.
XX	PF	22-DEC-1989; 89EP-00313573.
XX	PR	23-DEC-1988; 88GB-00030124.
XX	PR	01-FEB-1989; 89GB-00002178.
XX	PA	(CONN-) CONNAUGHT LAB LTD.
XX	PA	(UNIW) WASHINGTON UNIV ST LOUIS.
XX	PI	Munson RS, Tolani RW, Chong P, Fahim R, Mcverry P, Klein M;
XX	DR	WPI; 1990-225607/30.
XX	DR	N-PSDB; AAQ06320.
XX	PT	Gene coding for protein P2 of Haemophilus influenzae type-B - used for developing vaccines for protection against disease caused by the organism.
XX	PS	Claim 1; Page 8; 15pp; English.
XX	CC	Peptide may be used as a vaccine to the disease caused by HI type b, as carrier for conjugation to oligosaccharide derived from Haemophilus.
XX	CC	Protein may be used with other haptens as T-cell dependant antigen and carrier. (Updated on 25-MAR-2003 to correct PA field.)
XX	SQ	Sequence 341 AA;
	Query Match	37.5%; Score 6; DB 2; Length 341;
	Best Local Similarity	100.0%; Pred. No. 2.5e+02;
	Matches	6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	7 LPNKA 12	
Db	166 LPNKA 171	
	RESULT 7	
AAR34584	ID	AAR34584 standard; protein; 411 AA.
XX	AC	AAR34584;
XX	DT	25-MAR-2003 (revised)
XX	DT	14-SEP-1993 (first entry)
XX	DE	Mutant human prourokinase.
XX	KW	pUK; increased half life; improved fibrin affinity.
XX	OS	Homo sapiens.
XX	PN	PDE41952-A1

XX	PD	19-MAY-1993.
XX	PF	06-OCT-1992; 92EP-00117000.
XX	PR	07-OCT-1991; 91JP-00289257.
XX	PA	(GREC) GREEN CROSS CORP.
XX	PI	Tanabe T, Morita M, Hirose M, Amatsuji Y;
XX	DR	WPI; 1993-160551/20.
XX	DR	N-PSDB; AAQ41450.
XX	PT	New human pro-uokinase mutants with thrombolytic activity - have a neutral aminoacid in the epidermal growth factor region replaced with a basic aminoacid, or an acid residue replaced with a non-acidic residue.
XX	PS	Claim 1; Page 17-20; 38pp; English.
XX	CC	The sequence is that of a mutant human prourokinase, in which a neutral amino acid in the epidermal growth region has been replaced with a basic amino acid, or an acidic amino acid has been replaced by a non-acidic amino acid. Preferred replacements are 16Gly->Lys, 38Gly->Lys and 45Asp->Asn. The mutant has an increased half-life in blood as compared to the prior art mutant with a deleted RGF region. It has improved affinity for fibrin, and has other features the same as human prourokinase. (Updated on 25-MAR-2003 to correct PN field.)
XX	SQ	Sequence 411 AA;
	Query Match	37.5%; Score 6; DB 2; Length 411;
	Best Local Similarity	100.0%; Pred. No. 2.9e+02;
	Matches	6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	11 KASDDT 16	
Db	61 KASDDT 66	
	RESULT 8	
AAV51378	ID	AAV51378 standard; protein; 465 AA.
XX	AC	RAY51378;
XX	DT	04-MAY-2000 (first entry)
XX	DE	C. elegans SPE-4 protein.
XX	XW	SPE-4; immunogen; diagnosis; nematode; Alzheimer's disease; brain; muscle; peripheral blood cell; neuroprotectant; nootropic.
XX	OS	Caenorhabditis elegans.
XX	PN	US6019974-A.
XX	PD	01-FEB-2000.
XX	PF	24-JAN-1997; 97US-00788231.
XX	PR	26-JAN-1996; 96US-0010672P.
XX	PA	(UYEM-) UNIV EMORY.
XX	PI	L'hernault SW;
XX	DR	WPI; 2000-146863/13.
XX	DR	N-PSDB; AAZ88555.
XX	PT	Immunogenic composition comprising an antigenic component, useful for diagnosing Alzheimer's disease.
XX	PN	PDE41952-A1

Db	137 EFNSLP 142	
	RESULT 6	
AAR07043	ID	AAR07043 standard; protein; 341 AA.
XX	AC	AAR07043;
XX	DT	25-MAR-2003 (revised)
XX	DT	04-DEC-1990 (first entry)
XX	DE	P3 gene product of Haemophilus strain 8358 (OMP subtype 6U).
XX	DE	P3 gene; strain 1H; influenza; vaccine; ds.
XX	KW	Haemophilus influenzae.
XX	OS	Haemophilus influenzae.
XX	PN	EP378929-A.
XX	PD	25-JUL-1990.
XX	PF	22-DEC-1989; 89EP-00313573.
XX	PR	23-DEC-1988; 88GB-00030124.
XX	PR	01-FEB-1989; 89GB-00002178.
XX	PA	(CONN-) CONNAUGHT LAB LTD.
XX	PA	(UNIW) WASHINGTON UNIV ST LOUIS.
XX	PI	Munson RS, Tolian RW, Chong P, Fahim R, Mcverry P, Klein M;
XX	DR	WPI; 1990-225607/30.
XX	DR	N-PSDB; AAQ06320.
XX	PT	Gene coding for protein P2 of Haemophilus influenzae type-B - used for developing vaccines for protection against disease caused by the organism.
XX	PS	Claim 1; Page 8; 15pp; English.
XX	CC	Peptide may be used as a vaccine to the disease caused by HI type b, as carrier for conjugation to oligosaccharide derived from Haemophilus.
XX	CC	Protein may be used with other haptens as T-cell dependant antigen and carrier. (Updated on 25-MAR-2003 to correct PA field.)
XX	SQ	Sequence 341 AA;
	Query Match	37.5%; Score 6; DB 2; Length 341;
	Best Local Similarity	100.0%; Pred. No. 2.5e+02;
	Matches	6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	7 LPNKA 12	
Db	166 LPNKA 171	
	RESULT 7	
AAR34584	ID	AAR34584 standard; protein; 411 AA.
XX	AC	AAR34584;
XX	DT	25-MAR-2003 (revised)
XX	DT	14-SEP-1993 (first entry)
XX	DE	Mutant human prourokinase.
XX	KW	pUK; increased half life; improved fibrin affinity.
XX	OS	Homo sapiens.
XX	PN	PDE41952-A1

PS Example 1; Col 23-26; 23pp; English.

XX This invention describes a novel immunogenic composition comprising at
CC least 1 antigenic component selected from an antigenic peptide (I)
CC (linked to a carrier) or a multiantigenic peptide (II). The composition
CC is useful for diagnosing or monitoring SPE-4 related protein profiles of
CC nematodes and/or Alzheimer's disease patients, either in postmortem
CC tissue, or from other tissue samples, where the tissue is from the brain,
CC muscle or peripheral blood cells. The immunogenic composition can be used
CC to diagnose Alzheimer's noninvasively and has neuroprotective and
CC nootropic activity. This sequence represents the SPE-4 protein isolated
CC from *Caenorhabditis elegans* which is used in the method of the invention
XX

SQ Sequence 465 AA;

Query Match 37.5%; Score 6; DB 3; Length 465;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NSLPNN 10
DB 366 NSLPNN 371
|||||

RESULT 9
AAU08491
ID AAU08491 standard; protein; 573 AA.

XX AC AAU08491;

XX 17-DEC-2001 (first entry)

DE Mouse VMGLOM long form polypeptide.

XX Human; VMGLOM; glomulin; venous malformation glomangioma; cancer; mouse;
XX vascular smooth muscle cell; varicosity; cardiopathy; cardiomyopathy;
KW cerebral disorder; gene therapy; vasotrophic; cerebroprotective;
XX cytostatic; immunomodulator.

XX Mus sp.

XX WO200160856-A2.

XX 23-AUG-2001.

XX 16-FEB-2001; 2001WO-EP001760.

XX 16-FEB-2000; 2000EP-00870022.

XX 10-APR-2000; 2000US-0195777P.

XX 22-DEC-2000; 2000EP-00870320.

XX (UYLO-) UNIV CATHOLIQUE LOUVAIN.

XX Vikkula M;

XX WPI; 2001-557643/62.

XX N-PSDB; AAS13483.

XX New VMGLOM genes and polypeptides, useful in gene therapy or for

PT preventing, treating or alleviating disorders with vascular component,

PT e.g. varicosities, cardiopathies, cerebral disorders or cancer.

XX Claim 5; Fig 12; 157pp; English.

XX The present invention relates to the isolation of novel human and mouse
CC VMGLOM polypeptides (long form and short form), and the nucleic acid
CC molecules encoding them. VMGLOMs (also referred to as glomulins) are a
CC subtype of venous malformations (VMs) called glomangiomas. In humans,
CC VMGLOM has been mapped to chromosome 1p21-22. VMGLOMs and the nucleic
CC acids encoding for them are useful as a medicament or for inventing
CC into a diagnostic kit. Such medicaments are useful for preventing,
CC treating or alleviating disorders with a vascular component, particularly
CC where alteration of vascular smooth muscle cell phenotype is needed, e.g.

CC varicosities, cardiopathies or cardiomyopathies, cerebral disorders and
CC cancer. The nucleic acids are also useful in gene therapy. The present
CC sequence represents mouse VMGLOM "long form" polypeptide
XX

SQ Sequence 573 AA;

Query Match 37.5%; Score 6; DB 4; Length 573;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LPNNKA 12
DB 135 LPNNKA 140
|||||

RESULT 10
ADE60772
ID ADE60772 standard; protein; 596 AA.

XX AC ADE60772;

XX 29-JAN-2004 (first entry)

XX Rat Protein AAH03446, SEQ ID NO 6684.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX *Rattus norvegicus*.

XX WO2003016475-A2.

XX 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

XX 01-NOV-2001; 2001US-0346382P.

XX 26-NOV-2001; 2001US-0333347P.

XX (GEHO) GEN HOSPITAL CORP.

XX (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

XX GENBANK; AAH03446.

XX New composition comprising two or more isolated polypeptides, useful for

PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating

CC pain (e.g. spinal segmental nerve injury (Chung)), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 596 AA;

Query Match 37.5%; Score 6; DB 7; Length 596;
 Best Local Similarity 100.0%; Pred. No. 4e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LPNKA 12
 |||||
 Db 135 LPNKA 140

RESULT 11
 ABU08103
 ID ABU08103 standard; protein; 642 AA.

XX AC ABU08103;

XX 10-MAY-2003 (first entry)

XX Human kinase and phosphatase protein, KPP-4, INCYTE No. 90044205CD1.

XX Human; enzyme; kinase and phosphatase; KPP; cancer; cirrhosis;
 XX cell proliferative disorder; arteriosclerosis; atherosclerosis;
 XX hepatitis; paroxysmal nocturnal haemoglobinuria; polycythaemia vera;
 XX psoiasis; primary chromocytocytopenia; developmental disorder;
 XX renal tubular acidosis; anaemia; mental retardation; AIDS; epilepsy;
 XX neurological disorder; Alzheimer's disease; Parkinson's disease;
 XX autoimmune disorder; inflammatory disorder; allergy; asthma;
 XX acquired immunodeficiency syndrome; autoimmune thyroiditis;
 XX contact dermatitis; Crohn's disease; diabetes mellitus;
 XX glomerulonephritis; Goodpasture's syndrome; gout; Graves' disease;
 XX Hashimoto's thyroiditis; irritable bowel syndrome; multiple sclerosis;
 XX osteoarthritis; osteoporosis; pancreatitis; Reiter's syndrome;
 XX rheumatoid arthritis; Sjogren's syndrome; uveitis; infection.

XX Homo sapiens.

XX WO2003012065-A2.

XX 13-FEB-2003.

XX 01-AUG-2002; 2002WO-US024521.

XX 02-AUG-2001; 2001US-0309627P.

XX 07-AUG-2001; 2001US-0310933P.

XX 09-AUG-2001; 2001US-0311323P.

XX 07-SEP-2001; 2001US-0317820P.

XX 14-SEP-2001; 2001US-0322264P.

XX 28-SEP-2001; 2001US-0326098P.

XX 19-DEC-2001; 2001US-0343007P.

XX 15-MAR-2002; 2002US-0364494P.

XX 24-APR-2002; 2002US-0375539P.

XX (INCY-) INCYTE GENOMICS INC.

XX Baughn MR, Yue H, Wallia NK, He A, Au-Young JK, Lee SY;

XX Gietzen KJ, Lal PG, Elliott VS, Ison CH, Yang J, Lee EA, Li JX;

XX Emerling BM, Richardson TW, Warren BA, Hafalia AJA, Marquis JP;

XX WPI; 2003-239519/23.

PT cancer or hepatitis.

XX Claim 1; Page 161-162; 199pp; English.

XX The invention relates to an isolated polypeptide, which is a human kinase
 CC and phosphatase, KPP (KPP-1 to KPP-18). Also included are the encoding
 CC polynucleotides KPP NA, recombinant polynucleotide comprising a promoter
 CC sequence operably linked to KPP NA, a cell transformed with the
 CC recombinant polynucleotide, a transgenic organism comprising the
 CC recombinant polynucleotide; an anti-KPP antibody, screening for
 CC antiagonists of KPP, screening for compounds which bind to or alter the
 CC activity or expression of KPP, microarray where at least one element is
 CC KPP NA, generating an expression profile of a sample containing
 CC polynucleotides and an array comprising different nucleotide molecules
 CC affixed in distinct physical locations on a solid substrate (where at
 CC least one of the nucleotide molecules comprises a first oligonucleotide
 CC or polynucleotide sequence specifically hybridisable with at least 30
 CC contiguous nucleotides of a target KPP NA). The kinases and phosphatases
 CC (KPP) polypeptides, polynucleotides, agonists and antagonists are useful
 CC for diagnosing, treating or preventing disorders associated with aberrant
 CC expression of KPP, particularly cell proliferative disorders (e.g.
 CC arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, paroxysmal
 CC nocturnal haemoglobinuria, polycythaemia vera, psoriasis, primary
 CC thrombocytopenia or cancer), developmental disorders (renal tubular
 CC acidosis, anaemia or mental retardation), neurological disorders (e.g.
 CC Alzheimer's disease, Parkinson's disease or epilepsy), autoimmune/
 CC inflammatory disorders (e.g. AIDS, acquired immunodeficiency syndrome,
 CC allergies, asthma, autoimmune thyroiditis, contact dermatitis, Crohn's
 CC disease, diabetes mellitus, glomerulonephritis, Goodpasture's syndrome,
 CC gout, Graves' disease, Hashimoto's thyroiditis, irritable bowel syndrome,
 CC multiple sclerosis, osteoarthritis, osteoporosis, pancreatitis, Reiter's
 CC syndrome, rheumatoid arthritis, Sjogren's syndrome, uveitis), or viral,
 CC bacterial, fungal, parasitic, protozoan or helminthic infections. They
 CC are also useful in the assessment of the effects of exogenous compounds
 CC on the expression of nucleic acid and aa sequences of proteins associated
 CC with KPP. The polynucleotides encoding KPP are useful for creating
 CC transgenic animals to model human disease. The present sequence
 CC represents a KPP protein of the invention

XX Sequence 642 AA;

Query Match 37.5%; Score 6; DB 6; Length 642;

Best Local Similarity 100.0%; Pred. No. 4.3e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8
 |||||
 Db 80 EFNSLP 85

RESULT 12

ABU04245

ID ABU04245 standard; protein; 642 AA.

XX AC ABU04245;

XX 29-JAN-2003 (first entry)

XX Human expressed protein tag (EPT) #911.

XX Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KW protease; protease inhibitor; transporter; cytoskeletal protein;
 KW receptor; transcription factor; cancer; MHC;
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 XX adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

OS Homo sapiens.

XX WO200278524-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US009671.

XX New human kinases and phosphatases and polynucleotides, useful for

XX diagnosing, treating or preventing autoimmune or inflammatory disorders

XX (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,

XX 28-MAR-2001; 2001US-0279495P.
 PR 21-MAY-2001; 2001US-0292544P.
 PR 08-AUG-2001; 2001US-0310801P.
 PR 01-OCT-2001; 2001US-0326370P.
 PR 04-DEC-2001; 2001US-0336780P.
 PR 20-FEB-2002; 2002US-0358985P.
 XX (ZYCO-) ZYCOS INC.

XX Chicz RM, Tomlinson AJ, Urban RG;
 PI WPI; 2003-040607/03.

XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
 PT leukemia.

XX Example 2; SEQ ID NO 911; 134pp; English.

XX The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor. The
 CC polypeptide is useful as an immunogenic composition for eliciting in a
 CC mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to this
 CC polypeptide, is useful for treating cancer. The polypeptide is also
 CC useful for identifying compounds that binds to a naturally processed
 CC class I or class II MHC-binding polypeptide. The polypeptides and
 CC polynucleotides are particularly useful for treating or preventing
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC lymphoma or leukaemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational
 CC profiling. Note: This sequence does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 642 AA;

Query Match 37.5%; Score 6; DB 6; Length 642;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8
 |||||
 Db 80 EFNSLP 85

RESULT 13
 ABU04254
 ID ABU04254 standard; protein; 642 AA.

XX AC ABU04254;
 XX 29-JAN-2003 (first entry)

XX Human expressed protein tag (EPT) #920.

XX Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KW protease; protease inhibitor; transporter; cytoskeletal protein;
 KW receptor; transcription factor; cancer; MHC;
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX Homo sapiens.
 XX WO200278524-A2.
 XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US009671.

XX 28-MAR-2001; 2001US-0279495P.
 PR 21-MAY-2001; 2001US-0292544P.
 PR 08-AUG-2001; 2001US-0310801P.
 PR 01-OCT-2001; 2001US-0326370P.
 PR 04-DEC-2001; 2001US-0336780P.
 PR 20-FEB-2002; 2002US-0358985P.
 XX (ZYCO-) ZYCOS INC.

XX Chicz RM, Tomlinson AJ, Urban RG;
 PI WPI; 2003-040607/03.

XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
 PT leukemia.

XX Example 2; SEQ ID NO 920; 134pp; English.

XX The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor. The
 CC polypeptide is useful as an immunogenic composition for eliciting in a
 CC mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to this
 CC polypeptide, is useful for treating cancer. The polypeptide is also
 CC useful for identifying compounds that binds to a naturally processed
 CC class I or class II MHC-binding polypeptide. The polypeptides and
 CC polynucleotides are particularly useful for treating or preventing
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC lymphoma or leukaemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational
 CC profiling. Note: This sequence does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 642 AA;

Query Match 37.5%; Score 6; DB 6; Length 642;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8
 |||||
 Db 80 EFNSLP 85

RESULT 14
 AAW27226
 ID AAW27226 standard; protein; 699 AA.

XX AC AAW27226;

XX 25-MAR-2003 (revised)
 DT 19-DEC-1997 (first entry)

XX Mouse protein tyrosine phosphatase PTPEpsilon.

XX Protein tyrosine phosphatase' PTP-OB; PTPEpsilon; osteoblast;
 KW recombinant protein; growth; differentiation; brain; mouse.

XX Mus musculus.

XX Key Location/Qualifiers
 FT Peptide 1..17
 FT /label= Signal
 FT Region 48..67
 FT /label= Transmembrane
 FT Region 154..400
 FT /note= "Conserved PTP region"

FT Region 452..699
 FT /note= "Conserved PTP region"
 XX
 XX US5658756-A.
 XX
 XX PN
 XX PD 19-AUG-1997.
 XX
 XX PF 01-DEC-1994; 94US-00348006.
 XX
 XX PR 14-SEP-1993; 93US-00122032.
 XX
 XX PA (MERI) MERCK & CO INC.
 XX
 XX PI Schmidt A, Rodan GA, Rutledge SJ;
 XX
 XX DR WPI; 1997-424232/39.
 XX
 XX PT DNA encoding protein tyrosine phosphatase PTP-OB - isolated from human
 XX osteoblasts and useful for production of recombinant PTP-OB.
 XX
 XX PS Disclosure; Col 39-42; 34pp; English.
 XX
 XX CC The present sequence represents mouse protein tyrosine phosphatase
 XX (PTPepsilon) protein. The DNA encoding this protein is useful for the
 XX production of the recombinant protein, which is a protein tyrosine
 XX phosphatase which may be involved in the growth and differentiation of
 XX osteoblasts and brain cells and is useful for identifying compounds that
 XX modulate PTP-OB activity and as a therapeutic agent for treating PTP-OB-
 XX related diseases. (Updated on 25-MAR-2003 to correct PF field.)
 XX
 XX SQ Sequence 699 AA;

Query Match 37.5%; Score 6; DB 2; Length 699;
 Best Local Similarity 100.0%; Pred.No. 4.6e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 EFNSLP 8
 |||||
 Db 137 EFNSLP 142

RESULT 15
 AAW94028
 ID AAW94028 standard; protein; 699 AA.
 XX
 XX AC AAW94028;
 XX
 XX DT 01-APR-1999 (first entry)
 XX
 XX DE Mouse protein tyrosine phosphatase mPTPepsilon.
 XX
 XX KW Protein tyrosine phosphatase; PTP; PTP-OB; bone; brain; cancer;
 XX osteoporosis.
 XX
 XX OS Mus sp.

Key Location/Qualifiers
 FH Peptide 1..17
 FT /note= "signal peptide"
 FT Region 48..66
 FT /note= "transmembrane region"
 FT Region 154..399
 FT /note= "conserved PTP region"
 FT Region 452..699
 FT /note= "conserved PTP region"

XX US5866397-A.
 XX PN
 XX PD 02-FEB-1999.
 XX
 XX PF 14-FEB-1997; 97US-00800825.
 XX
 XX PR 14-SEP-1993; 93US-00122032.

PR 01-DEC-1994; 94US-00348006.
 XX
 XX PA (MERI) MERCK & CO INC.
 XX
 XX PI Rutledge SJ, Schmidt A, Rodan GA;
 XX
 XX DR WPI; 1999-141930/12.
 XX
 XX PT Protein tyrosine phosphatase denoted PTP-OB - useful for drug screening.
 XX
 XX PS Disclosure; Fig 11; 34pp; English.
 XX
 XX CC The invention relates to a human protein tyrosine phosphatase (PTP)
 XX denoted as PTP-OB, produced by bone and brain cells. A recombinant host
 XX cell transfected or transformed with a nucleic acid vector comprising the
 XX nucleic acid can be used for the production of the PTP-OB protein. The
 XX protein can be used to screen for modulators of PTP-OB activity, which
 XX might be useful for treating e.g. osteoporosis and cancer. The present
 XX sequence represents a mouse mPTPepsilon polypeptide sequence
 XX
 XX SQ Sequence 699 AA;

Query Match 37.5%; Score 6; DB 2; Length 699;
 Best Local Similarity 100.0%; Pred.No. 4.6e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 EFNSLP 8
 |||||
 Db 137 EFNSLP 142

Search completed: April 23, 2004, 14:51:40
 Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:49:06 ; Search time 14 seconds
(without alignments)
59.001 Million cell updates/sec

Title: US-09-528-682-4_COPY_64_79
Sequence: 1 VRENSLPNNKASDST 16

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	37.5	27	1	US-08-475-989-33
2	6	37.5	27	2	US-08-475-985-33
3	6	37.5	27	3	US-08-256-839-33
4	6	37.5	465	1	US-08-788-231A-2
5	6	37.5	699	1	US-08-348-006B-7
6	6	37.5	699	2	US-08-800-825A-7
7	6	37.5	699	3	US-09-158-657-7
8	6	37.5	888	2	US-08-861-464-6
9	6	37.5	888	3	US-08-396-001-6
10	6	37.5	888	3	US-09-323-433A-6
11	6	37.5	1161	1	US-08-173-497-2
12	6	37.5	1161	1	US-08-286-889-2
13	6	37.5	1161	1	US-08-485-618-2
14	6	37.5	1161	1	US-08-485-618-99
15	6	37.5	1161	1	US-08-362-652-2
16	6	37.5	1161	2	US-08-605-672-2
17	6	37.5	1161	2	US-08-605-672-99
18	6	37.5	1161	2	US-08-482-293A-2
19	6	37.5	1161	2	US-08-482-293A-99
20	6	37.5	1161	2	US-08-943-363-2
21	6	37.5	1161	2	US-08-943-363-99
22	6	37.5	1161	3	US-09-193-043-2
23	6	37.5	1161	3	US-09-193-043-99
24	6	37.5	1161	4	US-09-688-307A-2
25	6	37.5	1161	4	US-09-688-307A-99
26	6	37.5	1161	4	US-09-350-259-2
27	6	37.5	1161	4	US-09-350-259-99

Sequence 223, Appl
Sequence 13, Appl
Sequence 25, Appl
Sequence 13, Appl
Sequence 13, Appl
Sequence 13, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 5281, Ap
Sequence 7088, Ap
Sequence 5637, Ap
Sequence 7368, Ap
Sequence 7538, Ap
Sequence 4904, Ap
Sequence 3283, Ap
Sequence 149, Ap
Sequence 384, Ap
Sequence 528, Ap
Sequence 9803, Ap

28 5 31.2 32 4 US-09-227-357-223
29 5 31.2 35 3 US-08-513-968-13
30 5 31.2 35 3 US-08-944-483-25
31 5 31.2 66 2 US-08-588-258B-13
32 5 31.2 66 3 US-08-460-505-13
33 5 31.2 66 5 PCT-US96-08295-13
34 5 31.2 71 4 US-09-679-409-23
35 5 31.2 80 4 US-09-543-681A-5281
36 5 31.2 80 4 US-09-621-976-7088
37 5 31.2 115 4 US-09-543-681A-5637
38 5 31.2 115 4 US-09-543-681A-7368
39 5 31.2 116 4 US-09-543-681A-7538
40 5 31.2 123 4 US-09-134-000C-4904
41 5 31.2 126 4 US-09-732-210-149
42 5 31.2 130 4 US-08-936-165A-384
43 5 31.2 130 4 US-09-198-452A-528
44 5 31.2 142 4 US-09-489-039A-9803
45 5 31.2

ALIGNMENTS

RESULT 1
US-08-475-989-33
; Sequence 33, Application US/08475989
; Patent No. 5679352
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: KANDIL, Ali
; APPLICANT: SIA, Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: Synthetic Haemophilus Influenzae
; TITLE OF INVENTION: Conjugate Vaccine
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,989
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,839
; FILING DATE: 03-FEB-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA93/00041
; FILING DATE: 03-FEB-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9202219.3
; FILING DATE: 03-FEB-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-505 MIS:V9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-475-989-33

Query Match 37.5%; Score 6; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 LPNKA 12
Db 19 LPNKA 24

RESULT 2
US-08-475-985-33
Sequence 33, Application US/08475985
Patent No. 5972349
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: KANDIL, Ali
APPLICANT: SIA, Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: Synthetic Haemophilus Influenzae
TITLE OF INVENTION: Conjugate Vaccine
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,985
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: US 08/256,839
FILING DATE: 03-FEB-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/CA93/00041
FILING DATE: 03-FEB-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9202219.3
FILING DATE: 03-FEB-1992
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-506 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-475-985-33

Query Match 37.5%; Score 6; DB 2; Length 27;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 LPNKA 12
Db 19 LPNKA 24

RESULT 3
US-08-256-839-33
Sequence 33, Application US/08256839
Patent No. 6018019
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: KANDIL, Ali
APPLICANT: SIA, Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: Synthetic Haemophilus Influenzae
TITLE OF INVENTION: Conjugate Vaccine
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,839
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-373 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-256-839-33

Query Match 37.5%; Score 6; DB 3; Length 27;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 LPNKA 12
Db 19 LPNKA 24

RESULT 4
US-08-788-231A-2
Sequence 2, Application US/08788231A
Patent No. 6019974
GENERAL INFORMATION:
APPLICANT: L'Hernault, Steven W.
TITLE OF INVENTION: SPE-4 RELATED PEPTIDES, ANTIBODIES AND
TITLE OF INVENTION: METHODS
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/788,231A
FILING DATE: 24-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/010,672
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 60-95
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 465 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: YES
US-08-788-231A-2

Query Match 37.5%; Score 6; DB 3; Length 465;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NSLPNN 10
DB 366 NSLPNN 371

RESULT 5
US-08-348-006B-7
Sequence 7, Application US/08348006B
Patent No. 5658756
GENERAL INFORMATION:
APPLICANT: RODAN, GIDEON A.
APPLICANT: SCHMIDT, AZRIEL
APPLICANT: RUTLEDGE, SU JANE
TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN
TITLE OF INVENTION: TYROSINE PHOSPHATASE
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: J. MARK HAND
STREET: 126 E. LINCOLN AVE., P.O. BOX 2000
CITY: RAHWAY
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/348,006B
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/122,032
FILING DATE: 14-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: HAND, J., MARK
REGISTRATION NUMBER: 36,545
REFERENCE/DOCKET NUMBER: 189921A
TELECOMMUNICATION INFORMATION:

TELEPHONE: 908-594-3905
TELEFAX: 908-594-4720
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 699 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-348-006B-7

Query Match 37.5%; Score 6; DB 1; Length 699;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8
DB 137 EFNSLP 142

RESULT 6
US-08-800-825A-7
Sequence 7, Application US/08800825A
Patent No. 5866397
GENERAL INFORMATION:
APPLICANT: RODAN, GIDEON A.
APPLICANT: SCHMIDT, AZRIEL
APPLICANT: RUTLEDGE, SU JANE
TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN
TITLE OF INVENTION: TYROSINE PHOSPHATASE
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: J. MARK HAND - MERCK & CO., INC.
STREET: 126 E. LINCOLN AVE., P.O. BOX 2000
CITY: RAHWAY
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/800,825A
FILING DATE: 14-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HAND, J. MARK
REGISTRATION NUMBER: 36,545
REFERENCE/DOCKET NUMBER: 18992DA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 732-594-3905
TELEFAX: 732-594-4720
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 699 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-800-825A-7

Query Match 37.5%; Score 6; DB 2; Length 699;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8
DB 137 EFNSLP 142

RESULT 7

US-09-158-657-7
; Sequence 7, Application US/09158657
; Patent No. 6214564
; GENERAL INFORMATION:
; APPLICANT: RODAN, GIDEON A.
; APPLICANT: SCHMIDT, AZRIEL
; APPLICANT: RUTLEDGE, SU JANE
; TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: J. MARK HAND - MERCK & CO., INC.
; STREET: 126 E. LINCOLN AVE., P.O. BOX 2000
; CITY: RAHWAY
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/158,657
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/800,825
; FILING DATE: 14-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: HAND, J. MARK
; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: 18992DA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-3905
; TELEFAX: 732-594-4720
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 699 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-158-657-7

Query Match 37.5%; Score 6; DB 3; Length 699;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8
Db 137 EFNSLP 142

RESULT 8
US-08-861-464-6
; Sequence 6, Application US/08861464
; Patent No. 5874210
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: Genes Determining Cellular Senescence
; TITLE OF INVENTION: in Yeast
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/861,464
FILING DATE: 22-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/396,001
FILING DATE: 28-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09351
FILING DATE: 15-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/107,408
FILING DATE: 15-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-6408A22
TELECOMMUNICATION INFORMATION:
TELEPHONE: 781-861-6240
TELEFAX: 781-861-9540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 888 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-861-464-6

Query Match 37.5%; Score 6; DB 2; Length 888;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 NFKASS 14
Db 195 NFKASS 200

RESULT 9
US-08-396-001-6
; Sequence 6, Application US/08396001
; Patent No. 5919618
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Claus, James
; APPLICANT: Cole, Francesca
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: Genes Determining Cellular Senescence in
; TITLE OF INVENTION: Yeast
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396,001
; FILING DATE: 28-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: MIT-6408A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 888 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-396-001-6

Query Match 37.5%; Score 6; DB 2; Length 888;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 NKKASS 14
|||||
Db 195 NKKASS 200

RESULT 10
US-09-323-433A-6
Sequence 6, Application US/09323433A
Patent No. 6218512
GENERAL INFORMATION:
APPLICANT: Guarente, Leonard P.
APPLICANT: Austriaco Jr., Nicanor
APPLICANT: Claus, James J.
APPLICANT: Cole, Francesca
APPLICANT: Kennedy, Brian
TITLE OF INVENTION: GENES DETERMINING CELLULAR SENESECE IN YEAST
FILE REFERENCE: 0050.1491-003
CURRENT APPLICATION NUMBER: US/09/323,433A
CURRENT FILING DATE: 1999-06-01
PRIOR APPLICATION NUMBER: US 08/396,001
PRIOR FILING DATE: 1995-02-28
PRIOR APPLICATION NUMBER: PCT/US94/09351
PRIOR FILING DATE: 1994-08-15
PRIOR APPLICATION NUMBER: US 08/107,408
PRIOR FILING DATE: 1993-08-16
NUMBER OF SEQ ID NOS: 48
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6
LENGTH: 888
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-323-433A-6

Query Match 37.5%; Score 6; DB 3; Length 888;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 NKKASS 14
|||||
Db 195 NKKASS 200

RESULT 11
US-08-173-497-2
Sequence 2, Application US/08173497
Patent No. 5437958
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van Der Vieren, Monica
TITLE OF INVENTION: No. 5437958el Human 2 Integrin Alpha
TITLE OF INVENTION: Subunit
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 S. Wacker Drive, 6300 Sears Tower
CITY: Chicago

STATE: Illinois
COUNTRY: USA
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/173,497
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5437958and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/31363
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-173-497-2

Query Match 37.5%; Score 6; DB 1; Length 1161;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 NKKASS 14
|||||
Db 901 NKKASS 906

RESULT 12
US-08-286-889-2
Sequence 2, Application US/08286889
Patent No. 5470953
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Mich
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/286,889
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: P38,659
REFERENCE/DOCKET NUMBER: 27866/32168
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856

```

; INFORMATION FOR SEQ ID NO: 2:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 1161 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
; US-08-286-889-2

```

```
Query Match      37.5%; Score 6; DB 1; Length 1161;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6: Conservative 0; Mismatches 0; Indels
```

```

RESULT 13
US-08-485-618-2
; Sequence 2, Application US/08485618
; Patent No. 5728533
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,618

```

Query Match 37.5%; Score 6; DB 1; Length 1161;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6: Conservative 0; Mismatches 0; Indels

Db 901 NNNKASS 906

RESULT 14
US-08-485-618-99
; Sequence 99, Application US/08485618
; Patent No. 5728533

```
Query Match      37.5%; Score 6; DB 1; Length 1161;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels
```

RESULT 15
US-08-362-652-2
Sequence 2, Application US/08362652
Patent No. 5766850
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

```

; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,652
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32391
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-362-652-2

```

```

Query Match          37.5%; Score 6; DB 1; Length 1161;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY          9 NNRKASS 14
Db          901 NNRKASS 906

```

```

Search completed: April 23, 2004, 14:55:50
Job time : 15 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:54:17 ; Search time 35.6667 Seconds
(without alignments)
124.026 Million cell updates/sec

Title: US-09-528-682-4_COPY_64_79
Perfect score: 16
Sequence: 1 VREFSLPNKASDPT 16

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1133595 seqs, 276475211 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1133595

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	37.5	47	12	US-10-424-599-217103
2	6	37.5	59	12	US-10-424-599-180196
3	6	37.5	88	12	US-10-424-599-223194
4	6	37.5	123	12	US-10-424-599-147842
5	6	37.5	130	12	US-10-424-599-169864
6	6	37.5	154	12	US-10-424-599-239802
7	6	37.5	177	12	US-10-424-599-143948
8	6	37.5	182	12	US-10-424-599-144401
9	6	37.5	255	12	US-10-424-599-157105
10	6	37.5	265	12	US-10-424-599-242001
11	6	37.5	265	15	US-10-369-493-12954
12	6	37.5	271	14	US-10-029-386-32518
13	6	37.5	289	12	US-10-424-599-226313
14	6	37.5	383	12	US-10-425-114-69376
15	6	37.5	404	12	US-10-425-114-71851

16	6	37.5	487	12	US-10-424-599-147435	Sequence 147435, A
17	6	37.5	537	12	US-10-425-114-38564	Sequence 38564, A
18	6	37.5	573	14	US-10-204-254A-6	Sequence 6, Appli
19	6	37.5	614	12	US-10-424-599-285253	Sequence 285253, A
20	6	37.5	659	15	US-10-366-547-67	Sequence 67, Appl
21	6	37.5	676	12	US-10-424-599-265583	Sequence 265583, A
22	6	37.5	699	15	US-10-366-547-65	Sequence 65, Appl
23	6	37.5	700	15	US-10-366-547-63	Sequence 63, Appl
24	6	37.5	728	12	US-10-425-114-37370	Sequence 37370, A
25	6	37.5	774	12	US-10-087-192-1017	Sequence 1017, Ap
26	6	37.5	888	9	US-09-826-752-6	Sequence 6, Appli
27	6	37.5	1011	14	US-10-032-585-7548	Sequence 7548, Ap
28	6	37.5	1161	9	US-09-350-259-2	Sequence 2, Appli
29	6	37.5	1161	9	US-09-350-259-99	Sequence 99, Appli
30	6	37.5	1161	10	US-09-891-943-2	Sequence 2, Appli
31	6	37.5	1161	10	US-09-891-943-99	Sequence 99, Appli
32	5	31.2	12	10	US-09-954-385-149	Sequence 149, App
33	5	31.2	17	14	US-10-239-313A-42	Sequence 42, Appl
34	5	31.2	30	9	US-09-864-761-38398	Sequence 38398, A
35	5	31.2	31	12	US-09-973-278-248	Sequence 248, App
36	5	31.2	32	10	US-09-983-802-223	Sequence 223, App
37	5	31.2	32	12	US-09-984-490-223	Sequence 223, App
38	5	31.2	36	9	US-09-764-898-231	Sequence 231, Appl
39	5	31.2	40	12	US-10-424-599-231711	Sequence 231711, A
40	5	31.2	41	10	US-09-764-891-4658	Sequence 4658, Ap
41	5	31.2	41	12	US-10-424-599-275672	Sequence 275672, A
42	5	31.2	43	9	US-09-864-761-45904	Sequence 45904, A
43	5	31.2	48	12	US-10-424-599-157833	Sequence 157833, A
44	5	31.2	48	12	US-10-424-599-239220	Sequence 239220, A
45	5	31.2	50	9	US-09-864-761-41605	Sequence 41605, A

ALIGNMENTS

RESULT 1

US-10-424-599-217103
; Sequence 217103, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 217103
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT MRT3847_38071C.1.pep
US-10-424-599-217103

Query Match 37.5% Score 6: DB 12; Length 47;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LPNKA 12
| | | | |
Db 10 LPNKA 15

RESULT 2

US-10-424-599-180196
; Sequence 180196, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 180196
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_133731C.1.pap
US-10-424-599-180196

Query Match 37.5%; Score 6; DB 12; Length 59;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 REFNSL 7
Db 30 REFNSL 35

RESULT 3
US-10-424-599-223194
; Sequence 223194, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 223194
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(88)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_43575C.1.pap
US-10-424-599-223194

Query Match 37.5%; Score 6; DB 12; Length 88;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FNSLPN 9
Db 51 FNSLPN 56

RESULT 4
US-10-424-599-147842
; Sequence 147842, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 147842
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_104523C.1.pap
US-10-424-599-147842

Query Match 37.5%; Score 6; DB 12; Length 123;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 NNKASS 14
Db 53 NNKASS 58

RESULT 5
US-10-424-599-169864
; Sequence 169864, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 169864
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(130)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_124400C.1.pap
US-10-424-599-169864

Query Match 37.5%; Score 6; DB 12; Length 130;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FNSLPN 9
Db 64 FNSLPN 69

RESULT 6
US-10-424-599-239802
; Sequence 239802, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 239802

; LENGTH: 154
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_58567C.1.pep
US-10-424-599-239802

Query Match 37.5%; Score 6; DB 12; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 KASDDT 16
| | | | |
Db 92 KASDDT 97

RESULT 7
US-10-424-599-143948
; Sequence 143948, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 143948
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(177)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_100999C.1.pep
US-10-424-599-143948

Query Match 37.5%; Score 6; DB 12; Length 177;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 SLPNNK 11
| | | | |
Db 49 SLPNNK 54

RESULT 8
US-10-424-599-144401
; Sequence 144401, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 144401
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure

; LOCATION: (1)..(182)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_101407C.1.pep
US-10-424-599-144401

Query Match 37.5%; Score 6; DB 12; Length 182;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 KASDDT 16
| | | | |
Db 156 KASDDT 161

RESULT 9
US-10-424-599-157105
; Sequence 157105, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 157105
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112888C.1.pep
US-10-424-599-157105

Query Match 37.5%; Score 6; DB 12; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 KASDDT 16
| | | | |
Db 203 KASDDT 208

RESULT 10
US-10-424-599-242001
; Sequence 242001, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 242001
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_60554C.1.pep
US-10-424-599-242001

Query Match 37.5%; Score 6; DB 12; Length 265;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FNSLPN 9
| | | | |
Db 26 FNSLPN 31

RESULT 11
US-10-369-493-12954
; Sequence 12954, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12954
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-369-493-12954

Query Match 37.5%; Score 6; DB 15; Length 265;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 SLPNNK 11
| | | | |
Db 38 SLPNNK 43

RESULT 12
US-10-029-386-32518
; Sequence 32518, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32518
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008762.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.81
; OTHER INFORMATION: SWISSPROT HIT: P16053, EVALUATE 3.00e-07
US-10-029-386-32518

Query Match 37.5%; Score 6; DB 14; Length 271;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8
| | | | |

Db 131 EFNSLP 136

RESULT 13
US-10-424-599-226313
; Sequence 226313, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 226313
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_46390C.1.pap
US-10-424-599-226313

Query Match 37.5%; Score 6; DB 12; Length 289;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 NKASSD 15
| | | | |
Db 185 NKASSD 190

RESULT 14
US-10-425-114-69376
; Sequence 69376, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 69376
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMROPIC011A12_FLI.pap
US-10-425-114-69376

Query Match 37.5%; Score 6; DB 12; Length 383;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 SLPNNK 11
| | | | |
Db 154 SLPNNK 159

RESULT 15
US-10-425-114-71851
; Sequence 71851, Application US/10425114
; Publication No. US20040034888A1

GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 71851
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMELM017092C10_FL1.pep
US-10-425-114-71851

Query Match 37.5%; Score 6; DB 12; Length 404;
Best Local Similarity 100.0%; Pred.No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 FNSLEN 9
| | | | |
DB 280 FNSLEN 285

Search completed: April 23, 2004, 15:07:24
Job time : 35.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:51:47 ; Search time 173 Seconds
(without alignments)
90.271 Million cell updates/sec

Title: US-09-528-682-4_COPY_64_79
Perfect score: 16
Sequence: 1 VREFNSLPNNKASDST 16

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 6019581 seqs, 976053577 residues

Word size : 0
Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending Patents AA Main:*

- 1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US080_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
- 22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
- 23: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
- 24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
- 25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
- 26: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
- 27: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
- 28: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
- 29: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
- 30: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
- 31: /cgn2_6/ptodata/2/paa/US106_COMB.pep.*
- 32: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
- 33: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	16	100.0	110	19	US-09-528-682-4	Sequence 4, Appli
2	7	43.8	459	1	PCT-US03-30292-490	Sequence 490, App
3	7	43.8	459	31	US-10-666-642-490	Sequence 490, App
4	6	37.5	44	33	US-60-196-711-2080	Sequence 2080, Ap
5	6	37.5	47	30	US-10-424-599-217103	Sequence 217103,
6	6	37.5	59	30	US-10-424-599-180196	Sequence 180196,
7	6	37.5	60	19	US-09-513-999C-5917	Sequence 5917, Ap
8	6	37.5	75	33	US-60-196-710-3602	Sequence 3602, Ap
9	6	37.5	88	30	US-10-424-599-223194	Sequence 223194,
10	6	37.5	94	30	US-10-437-963-201277	Sequence 201277,
11	6	37.5	114	1	PCT-US03-29833-41	Sequence 41, Appl
12	6	37.5	116	1	PCT-US01-01307-523	Sequence 523, App
13	6	37.5	116	26	US-10-092-302-523	Sequence 523, App
14	6	37.5	123	30	US-10-424-599-147842	Sequence 147842,
15	6	37.5	124	19	US-09-538-092-98	Sequence 98, Appl
16	6	37.5	130	30	US-10-424-599-169864	Sequence 169864,
17	6	37.5	135	21	US-09-733-089-8937	Sequence 8937, Ap
18	6	37.5	135	23	US-09-816-660-8937	Sequence 8937, Ap
19	6	37.5	147	33	US-60-171-481-1758	Sequence 1758, Ap
20	6	37.5	154	30	US-10-424-599-239802	Sequence 239802,
21	6	37.5	154	33	US-60-177-571-4815	Sequence 4815, Ap
22	6	37.5	156	30	US-10-437-963-180356	Sequence 180356,
23	6	37.5	172	30	US-10-437-963-187184	Sequence 187184,
24	6	37.5	177	30	US-10-424-599-143948	Sequence 143948,
25	6	37.5	182	30	US-10-424-599-144401	Sequence 144401,
26	6	37.5	203	22	US-09-760-475-2208	Sequence 2208, Ap
27	6	37.5	203	28	US-10-227-425-2208	Sequence 2208, Ap
28	6	37.5	226	1	PCT-US01-01307-792	Sequence 792, App
29	6	37.5	226	26	US-10-092-302-792	Sequence 792, App
30	6	37.5	255	30	US-10-424-599-157105	Sequence 157105, A
31	6	37.5	265	29	US-10-369-493-12954	Sequence 12954, A
32	6	37.5	265	30	US-10-424-599-242001	Sequence 242001,
33	6	37.5	265	33	US-60-360-039-12954	Sequence 12954, A
34	6	37.5	271	26	US-10-029-386-32518	Sequence 32518, A
35	6	37.5	279	31	US-10-679-063-1681	Sequence 1681, Ap
36	6	37.5	289	22	US-09-791-537-118265	Sequence 118265,
37	6	37.5	289	27	US-10-155-881-18945	Sequence 18945, A
38	6	37.5	289	30	US-10-424-599-226313	Sequence 226313,
39	6	37.5	289	30	US-10-438-246-30738	Sequence 30738, A
40	6	37.5	290	27	US-10-179-131-7516	Sequence 7516, Ap
41	6	37.5	290	30	US-10-437-963-116475	Sequence 116475,
42	6	37.5	297	28	US-10-219-051B-2174	Sequence 2174, Ap
43	6	37.5	297	28	US-10-219-051B-10745	Sequence 10745, A
44	6	37.5	320	27	US-10-155-881-19846	Sequence 19846, A
45	6	37.5	321	30	US-10-437-963-194894	Sequence 194894,

ALIGNMENTS

RESULT 1

US-09-528-682-4
; Sequence 4, Application US/09528682
; GENERAL INFORMATION:
; APPLICANT: Pizzia, Mariagrazia
; APPLICANT: Giuliani, Marzia M
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: IMMUNOGENIC DETOXIFIED MUTANT E. COLI LT-A-TOXIN
; FILE REFERENCE: 2302-0342.10
; CURRENT APPLICATION NUMBER: US/09/528,682
; PRIOR FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: PCT/IB97/01440
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: 09/297,171
; PRIOR FILING DATE: 1999-04-27
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 4
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-528-682-4

Query Match 100.0%; Score 16; DB 19; Length 110;
Best Local Similarity 100.0%; Pred. No. 1.9e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VREFNSLPNNKASDST 16
DB 64 VREFNSLPNNKASDST 79
|||||

RESULT 2

PCT-US03-30292-490
; Sequence 490, Application PC/TUS0330292
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: JIANG, Cai-Zhong
; APPLICANT: HEARD, Jacqueline E
; APPLICANT: RATCLIFFE, Oliver
; APPLICANT: CREELMAN, Robert A
; APPLICANT: ADAM, Luc J
; APPLICANT: REUBER, T. Lynne
; APPLICANT: RIECHMANN, Jose Luis
; APPLICANT: HAAKE, Volker
; APPLICANT: DUBELL, Arnold N
; APPLICANT: KEDDIE, James S
; APPLICANT: SHERMAN, Bradley K
; TITLE OF INVENTION: Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MBI-0054
; CURRENT APPLICATION NUMBER: PCT/US03/30292
; CURRENT FILING DATE: 2003-09-22
; PRIOR APPLICATION NUMBER: 60/411,837
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: 60/434,166
; PRIOR FILING DATE: 2002-12-17
; PRIOR APPLICATION NUMBER: 60/465,809
; PRIOR FILING DATE: 2003-04-24
; NUMBER OF SEQ ID NOS: 2247
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 490
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
PCT-US03-30292-490

Query Match 43.8%; Score 7; DB 1; Length 459;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 SLPNNKA 12
DB 294 SLPNNKA 300
|||||

RESULT 3

US-10-666-642-490
; Sequence 490, Application US/10666642
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: JIANG, Cai-Zhong
; APPLICANT: HEARD, Jacqueline E
; APPLICANT: RATCLIFFE, Oliver
; APPLICANT: CREELMAN, Robert A
; APPLICANT: ADAM, Luc J
; APPLICANT: REUBER, T. Lynne
; APPLICANT: RIECHMANN, Jose Luis
; APPLICANT: HAAKE, Volker
; APPLICANT: DUBELL, Arnold N
; APPLICANT: KEDDIE, James S
; APPLICANT: SHERMAN, Bradley K
; TITLE OF INVENTION: Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MBI-0054
; CURRENT APPLICATION NUMBER: US/10/666,642
; CURRENT FILING DATE: 2003-09-18
; PRIOR APPLICATION NUMBER: 60/411,837

; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: 60/434,166
; PRIOR FILING DATE: 2002-12-17
; PRIOR APPLICATION NUMBER: 60/465,809
; NUMBER OF SEQ ID NOS: 2247
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 490
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-666-642-490

Query Match 43.8%; Score 7; DB 31; Length 459;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 SLPNNKA 12
DB 294 SLPNNKA 300
|||||

RESULT 4

US-60-196-711-2080
; Sequence 2080, Application US/60196711
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS OF THE
; TITLE OF INVENTION: SER/THR AND TYR FAMILY OF KINASES, NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING THESE HUMAN KINASE PROTEINS, AND USES THEREOF
; FILE REFERENCE: CLO00452
; CURRENT APPLICATION NUMBER: US/60/196,711
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 2378
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2080
; LENGTH: 44
; TYPE: PRT
; ORGANISM: HUMAN
US-60-196-711-2080

Query Match 37.5%; Score 6; DB 33; Length 44;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRENS 6
DB 31 VREFNS 36
|||||

RESULT 5

US-10-424-599-217103
; Sequence 217103, Application US/10424599
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 217103
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_38071C.1.pap
US-10-424-599-217103

Query Match 37.5%; Score 6; DB 30; Length 47;

Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LPNKA 12
|||||
Db 10 LPNKA 15

RESULT 6
US-10-424-599-180196
; Sequence 180196, Application US/10424599
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 180196
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_133731C.1.pep
US-10-424-599-180196

Query Match 37.5%; Score 6; DB 30; Length 59;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 REFNSL 7
|||||
Db 30 REFNSL 35

RESULT 7
US-09-513-999C-5917
; Sequence 5917, Application US/09513999C
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 5917
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-513-999C-5917

Query Match 37.5%; Score 6; DB 19; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 SLPNNK 11
|||||
Db 16 SLPNNK 21

RESULT 8
US-60-196-710-3602
; Sequence 3602, Application US/60196710
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000450
; CURRENT APPLICATION NUMBER: US/60/196,710
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 7166
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3602
; LENGTH: 75
; TYPE: PRT
; ORGANISM: HUMAN
US-60-196-710-3602

Query Match 37.5%; Score 6; DB 33; Length 75;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 NKKASS 14
|||||
Db 29 NKKASS 34

RESULT 9
US-10-424-599-223194
; Sequence 223194, Application US/10424599
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 223194
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(88)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_43575C.1.pep
US-10-424-599-223194

Query Match 37.5%; Score 6; DB 30; Length 88;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FNSLPN 9
|||||
Db 51 FNSLPN 56

RESULT 10
US-10-437-963-201277
; Sequence 201277, Application US/10437963
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 201277
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_96669C.1.pep
US-10-437-963-201277

Query Match 37.5%; Score 6; DB 30; Length 94;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 SLPNNK 11
|||
Db 42 SLPNNK 47

RESULT 11
PCT-US03-29833-41
; Sequence 41, Application PC/TUS0329833
; GENERAL INFORMATION:
; APPLICANT: THE GOVERNMENT OF THE UNITED STATES OF AMERICA AS
; APPLICANT: REPRESENTED BY THE SECRETARY OF THE DEPARTMENT OF HEALTH AND
; APPLICANT: HUMAN SERVICES
; APPLICANT: Valenzuela, Jesus G.
; APPLICANT: Ribeiro, Jose M.C.
; APPLICANT: Kamhawi, Shaden
; APPLICANT: Belkaid, Yasmine
; APPLICANT: Fischer, Laurent Bernard
; APPLICANT: Audonnet, Jean-Cristophe
; APPLICANT: Milward, Francis William
; TITLE OF INVENTION: P. ARIASI POLYPEPTIDES AND P. PERNICIOSUS POLYPEPTIDES AND
; FILE REFERENCE: 4239-66903
; CURRENT APPLICATION NUMBER: PCT/US03/29833
; CURRENT FILING DATE: 2003-09-18
; PRIOR APPLICATION NUMBER: US 60/425,852
; PRIOR FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/412,327
; PRIOR FILING DATE: 2002-09-19
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Phlebotomus ariasi
PCT-US03-29833-41

Query Match 37.5%; Score 6; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EFNSLP 8
|||
Db 38 EFNSLP 43

RESULT 12
PCT-US01-01307-523
; Sequence 523, Application PC/TUS0101307
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc., et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ18PCT
; CURRENT APPLICATION NUMBER: PCT/US01/01307
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1040
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 523
; LENGTH: 116
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (115)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US01-01307-523

Query Match 37.5%; Score 6; DB 1; Length 116;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 SLPNNK 11
|||
Db 35 SLPNNK 40

RESULT 13
US-10-092-302-523
; Sequence 523, Application US/10092302
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ18C1
; CURRENT APPLICATION NUMBER: US/10/092,302
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 1040
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 523
; LENGTH: 116
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (115)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-092-302-523

Query Match 37.5%; Score 6; DB 26; Length 116;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 SLPNNK 11
|||
Db 35 SLPNNK 40

RESULT 14
US-10-424-599-147842
; Sequence 147842, Application US/10424599
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 147842
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_104523C.1.pep
US-10-424-599-147842

Query Match 37.5%; Score 6; DB 30; Length 123;

Best Local Similarity 100.0%; Pred. No. 4.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 NKKASS 14
| | | | |
Db 53 NKKASS 58

RESULT 15
US-09-538-092-98
; Sequence 98, Application US/09538092
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormatter Version 0.9
; SEQ ID NO 98
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YDL071C
US-09-538-092-98

Query Match 37.5%; Score 6; DB 19; Length 124;
Best Local Similarity 100.0%; Pred. No. 4.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 SLPNNK 11
| | | | |
Db 40 SLPNNK 45

Search completed: April 23, 2004, 15:04:43
Job time : 175 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 23, 2004, 14:52:22 ; Search time 10 Seconds
(without alignments)
46.246 Million cell updates/sec

Title: US-09-528-682-4_COPY_64_79
Perfect score: 16
Sequence: 1 VREFNSLPNNKASDDT 16

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 202516 seqs, 28903602 residues

Word size: 0

Total number of hits satisfying chosen parameters: 202516

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Pending Patents AA New:
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	37.5	60	6 US-10-793-479-5917	Sequence 5917, Ap
2	6	37.5	61	6 US-10-767-701-59181	Sequence 59181, A
3	6	37.5	465	6 US-10-811-139-4	Sequence 4, Appli
4	6	37.5	700	6 US-60-556-903-259	Sequence 259, App
5	6	37.5	934	1 PCT-US03-24982A-117	Sequence 117, App
6	6	37.5	999	1 PCT-US04-07412-968	Sequence 968, App
7	5	31.2	12	6 US-10-111-983-12977	Sequence 12977, A
8	5	31.2	12	6 US-10-111-983-25070	Sequence 25070, A
9	5	31.2	13	6 US-10-111-983-7277	Sequence 7277, Ap
10	5	31.2	13	6 US-10-111-983-20606	Sequence 20606, A
11	5	31.2	13	6 US-10-111-983-36963	Sequence 36963, A
12	5	31.2	14	6 US-10-111-983-36953	Sequence 36953, A
13	5	31.2	15	6 US-10-111-983-7267	Sequence 7267, Ap
14	5	31.2	15	6 US-10-111-983-12967	Sequence 12967, A
15	5	31.2	15	6 US-10-111-983-20596	Sequence 20596, A
16	5	31.2	15	6 US-10-111-983-25060	Sequence 25060, A
17	5	31.2	31	6 US-10-100-683-6919	Sequence 6919, Ap
18	5	31.2	46	6 US-10-724-972A-4740	Sequence 4740, Ap
19	5	31.2	59	6 US-10-767-701-48541	Sequence 48541, A
20	5	31.2	62	6 US-10-793-479-7483	Sequence 7483, Ap
21	5	31.2	68	6 US-10-603-205-2	Sequence 2, Appli
22	5	31.2	73	6 US-10-767-701-49977	Sequence 49977, A
23	5	31.2	75	6 US-10-767-701-56848	Sequence 56848, A
24	5	31.2	79	6 US-10-767-701-39673	Sequence 39673, A
25	5	31.2	109	6 US-10-767-701-59603	Sequence 59603, A
26	5	31.2	112	6 US-10-767-701-40467	Sequence 40467, A

27 5 31.2 113 6 US-10-767-701-37840 Sequence 37840, A
28 5 31.2 123 6 US-10-724-972A-6414 Sequence 6414, Ap
29 5 31.2 125 6 US-10-767-701-50099 Sequence 50099, A
30 5 31.2 128 6 US-10-767-701-43249 Sequence 43249, A
31 5 31.2 133 6 US-10-767-701-31930 Sequence 31930, A
32 5 31.2 133 6 US-10-290-752-682 Sequence 682, App
33 5 31.2 143 1 PCT-US04-09202-957 Sequence 957, App
34 5 31.2 149 6 US-10-793-479-5821 Sequence 5821, App
35 5 31.2 152 6 US-10-491-823-275 Sequence 275, App
36 5 31.2 154 6 US-10-491-823-273 Sequence 273, App
37 5 31.2 159 6 US-10-767-701-35037 Sequence 35037, A
38 5 31.2 159 6 US-10-767-701-59644 Sequence 59644, A
39 5 31.2 159 6 US-10-491-823-254 Sequence 254, App
40 5 31.2 161 6 US-10-767-701-44843 Sequence 44843, A
41 5 31.2 161 6 US-10-100-683-10734 Sequence 10734, A
42 5 31.2 165 6 US-10-486-805-21 Sequence 21, Appli
43 5 31.2 175 6 US-10-767-701-59611 Sequence 59611, A
44 5 31.2 175 7 US-60-556-841-8184 Sequence 8184, Ap
45 5 31.2 177 6 US-10-491-823-259 Sequence 259, App

ALIGNMENTS

RESULT 1
US-10-793-479-5917
; Sequence 5917: Application US/10793479
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/10793.479
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: US/09/513,999
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 5917
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-793-479-5917

Query Match 37.5%; Score 6; DB 6; Length 60;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 SLPNNK 11
Db 16 SLPNNK 21

RESULT 2
US-10-767-701-59181
; Sequence 59181: Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Zhao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10767.701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 59181
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Sorghum bicolor

FEATURE:
OTHER INFORMATION: Clone ID: 7217487.pcp
US-10-767-701-59181

Query Match 37.5%; Score 6; DB 6; Length 61;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Gaps 0; Indels 0;

QY 9 NKKASS 14
|||||
Db 31 NKKASS 36

RESULT 3

US-10-811-199-4
; Sequence 4, Application US/10811199
; GENERAL INFORMATION:
; APPLICANT: Greenwald, Iva
; APPLICANT: Levitan, Diane
; TITLE OF INVENTION: IDENTIFICATION OF SEL 12 AND USES THEREOF
; FILE REFERENCE: 0575/48231-A-PCT-US
; CURRENT APPLICATION NUMBER: US/10/811,199
; CURRENT FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: US/09/043,944
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: PCT/US96/15727
; PRIOR FILING DATE: 1996-09-27
; PRIOR APPLICATION NUMBER: 60/004,387
; PRIOR FILING DATE: 1995-09-27
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 465
; TYPE: PRT
; ORGANISM: C. elegans
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Fig. 2A SPE-4
US-10-811-199-4

Query Match 37.5%; Score 6; DB 6; Length 465;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Gaps 0; Indels 0;

QY 5 NSLPNN 10
|||||
Db 366 NSLPNN 371

RESULT 4

US-60-556-903-259
; Sequence 259, Application US/60556903
; GENERAL INFORMATION:
; APPLICANT: Ford, Shirin K.
; APPLICANT: Perkins, Nancy-Anne A.
; APPLICANT: Jackson, Donald G.
; TITLE OF INVENTION: BIOMARKERS AND METHODS FOR DETERMINING SENSITIVITY TO EPIDERMAL
; FILE REFERENCE: GROWTH FACTOR RECEPTOR MODULATORS IN NON-SMALL CELL LUNG CANCER
; CURRENT APPLICATION NUMBER: US/60/556,903
; CURRENT FILING DATE: 2004-03-26
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 259
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-556-903-259

Query Match 37.5%; Score 6; DB 7; Length 700;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 0; Gaps 0; Indels 0;

QY 3 EFNSLP 8
|||||
Db 138 EFNSLP 143

RESULT 5

PCT-US03-24982A-117
; Sequence 117, Application PC/TUS0324982A
; GENERAL INFORMATION:
; APPLICANT: Syngenta Participations AG
; APPLICANT: Stam, Lynn
; APPLICANT: Kamdar, Kim
; APPLICANT: Spana, Eric
; APPLICANT: Bachmann, Jane
; TITLE OF INVENTION: Nucleic Acid Sequences from Drosophila Melanogaster that Encode P
; FILE REFERENCE: 70131WOPCT
; CURRENT APPLICATION NUMBER: PCT/US03/24982A
; PRIOR APPLICATION NUMBER: 2003-08-08
; PRIOR FILING DATE: 2002-10-29
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 117
; LENGTH: 934
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
PCT-US03-24982A-117

Query Match 37.5%; Score 6; DB 1; Length 934;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 6; Conservative 0; Mismatches 0; Gaps 0; Indels 0;

QY 9 NKKASS 14
|||||
Db 162 NKKASS 167

RESULT 6

PCT-US04-07412-968
; Sequence 968, Application PC/TUS0407412
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-rui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Ghosh, Malabika
; APPLICANT: Xue, Aidong J.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Dunrui
; APPLICANT: Goodrich, Kyle W.
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Gezhi
; APPLICANT: Wang, Zhiwei
; APPLICANT: Boyle, Bryan J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 822CIP/PCT
; CURRENT APPLICATION NUMBER: PCT/US04/07412
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: US 10/389,559
; PRIOR FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 60/365,264
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/340,187
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 10/296,115
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: PCT/US00/35017


```

; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 10/275,027
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1920
; SOFTWARE: pt_FL_genes Version 6.0
; SEQ ID NO 968
; LENGTH: 999
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-07412-968

```

```

Query Match          37.5%; Score 6; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 3 EFNSLP 8
DB 841 EFNSLP 846

```

```

RESULT 7
US-10-111-983-12977
; Sequence 12977, Application US/10111983
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: GALEOTTI Cesira
; APPLICANT: GRANDI Guido
; APPLICANT: MASIGNANI Vega
; APPLICANT: MORA Mariarosa
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: RAPPUOLI Rino
; APPLICANT: RATTI Giulio
; APPLICANT: SCARLATO Vincenzo
; APPLICANT: SCARSELLI Maria
; TITLE OF INVENTION: NEISSERIAL ANTIGENIC PEPTIDES
; FILE REFERENCE: 2300-1654 (PP01654.003)
; CURRENT APPLICATION NUMBER: US/10/111,983
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US-60/162616
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 37764
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 12977
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Neisseria
US-10-111-983-12977

```

```

Query Match          31.2%; Score 5; DB 6; Length 12;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 10 NKASS 14
DB 8 NKASS 12

```

```

RESULT 8
US-10-111-983-25070
; Sequence 25070, Application US/10111983
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: GALEOTTI Cesira
; APPLICANT: GRANDI Guido

```

```

; APPLICANT: MASIGNANI Vega
; APPLICANT: MORA Mariarosa
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: RAPPUOLI Rino
; APPLICANT: RATTI Giulio
; APPLICANT: SCARLATO Vincenzo
; APPLICANT: SCARSELLI Maria
; TITLE OF INVENTION: NEISSERIAL ANTIGENIC PEPTIDES
; FILE REFERENCE: 2300-1654 (PP01654.003)
; CURRENT APPLICATION NUMBER: US/10/111,983
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US-60/162616
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 37764
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 25070
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Neisseria
US-10-111-983-25070

```

```

Query Match          31.2%; Score 5; DB 6; Length 12;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 10 NKASS 14
DB 8 NKASS 12

```

```

RESULT 9
US-10-111-983-7277
; Sequence 7277, Application US/10111983
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: GALEOTTI Cesira
; APPLICANT: GRANDI Guido
; APPLICANT: MASIGNANI Vega
; APPLICANT: MORA Mariarosa
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: RAPPUOLI Rino
; APPLICANT: RATTI Giulio
; APPLICANT: SCARLATO Vincenzo
; APPLICANT: SCARSELLI Maria
; TITLE OF INVENTION: NEISSERIAL ANTIGENIC PEPTIDES
; FILE REFERENCE: 2300-1654 (PP01654.003)
; CURRENT APPLICATION NUMBER: US/10/111,983
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US-60/162616
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 37764
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 7277
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Neisseria
US-10-111-983-7277

```

```

Query Match          31.2%; Score 5; DB 6; Length 13;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 10 NKASS 14
DB 8 NKASS 12

```

```

RESULT 10
US-10-111-983-20606
; Sequence 20606, Application US/10111983
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: GALEOTTI Cesira

```

```

; APPLICANT: GRANDI Guido
; APPLICANT: MASIGNANI Vega
; APPLICANT: MORA Mariarosa
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: RAPPUOLI Rino
; APPLICANT: RATTI Giulio
; APPLICANT: SCARLETO Vincenzo
; APPLICANT: SCARSELLI Maria
; TITLE OF INVENTION: NEISSERIAL ANTIGENIC PEPTIDES
; FILE REFERENCE: 2300-1654 (PP01654.003)
; CURRENT APPLICATION NUMBER: US/10/111,983
; PRIOR FILING DATE: 2003-06-27
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 37764
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 20606
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Neisseria
US-10-111-983-20606

```

```

Query Match          31.2%; Score 5; DB 6; Length 13;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      10 NKASS 14
        |||||
Db       8 NKASS 12

```

```

RESULT 11
US-10-111-983-36963
; Sequence 36963, Application US/10111983
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: GALEOTTI Cesira
; APPLICANT: GRANDI Guido
; APPLICANT: MASIGNANI Vega
; APPLICANT: MORA Mariarosa
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: RAPPUOLI Rino
; APPLICANT: RATTI Giulio
; APPLICANT: SCARLETO Vincenzo
; APPLICANT: SCARSELLI Maria
; TITLE OF INVENTION: NEISSERIAL ANTIGENIC PEPTIDES
; FILE REFERENCE: 2300-1654 (PP01654.003)
; CURRENT APPLICATION NUMBER: US/10/111,983
; PRIOR FILING DATE: 2003-06-27
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 37764
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 36963
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Neisseria
US-10-111-983-36963

```

```

Query Match          31.2%; Score 5; DB 6; Length 13;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      10 NKASS 14
        |||||
Db       8 NKASS 12

```

```

RESULT 12
US-10-111-983-36953
; Sequence 36953, Application US/10111983
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa

```

```

; APPLICANT: GALEOTTI Cesira
; APPLICANT: GRANDI Guido
; APPLICANT: MASIGNANI Vega
; APPLICANT: MORA Mariarosa
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: RAPPUOLI Rino
; APPLICANT: RATTI Giulio
; APPLICANT: SCARLETO Vincenzo
; APPLICANT: SCARSELLI Maria
; TITLE OF INVENTION: NEISSERIAL ANTIGENIC PEPTIDES
; FILE REFERENCE: 2300-1654 (PP01654.003)
; CURRENT APPLICATION NUMBER: US/10/111,983
; PRIOR FILING DATE: 2003-06-27
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 37764
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 36953
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Neisseria
US-10-111-983-36953

```

```

Query Match          31.2%; Score 5; DB 6; Length 14;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      10 NKASS 14
        |||||
Db       8 NKASS 12

```

```

RESULT 13
US-10-111-983-7267
; Sequence 7267, Application US/10111983
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: GALEOTTI Cesira
; APPLICANT: GRANDI Guido
; APPLICANT: MASIGNANI Vega
; APPLICANT: MORA Mariarosa
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: RAPPUOLI Rino
; APPLICANT: RATTI Giulio
; APPLICANT: SCARLETO Vincenzo
; APPLICANT: SCARSELLI Maria
; TITLE OF INVENTION: NEISSERIAL ANTIGENIC PEPTIDES
; FILE REFERENCE: 2300-1654 (PP01654.003)
; CURRENT APPLICATION NUMBER: US/10/111,983
; PRIOR FILING DATE: 2003-06-27
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 37764
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 7267
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Neisseria
US-10-111-983-7267

```

```

Query Match          31.2%; Score 5; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      10 NKASS 14
        |||||
Db       8 NKASS 12

```

```

RESULT 14
US-10-111-983-12967
; Sequence 12967, Application US/10111983
; GENERAL INFORMATION:

```

```

; APPLICANT: CHIRON SpA
; APPLICANT: GALEOTTI Cesira
; APPLICANT: GRANDI Guido
; APPLICANT: MASIGNANI Vega
; APPLICANT: MORA Mariarosa
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: RAPPUOLI Rino
; APPLICANT: RATTI Giulio
; APPLICANT: SCARLATO Vincenzo
; APPLICANT: SCARSELLI Maria
; TITLE OF INVENTION: NEISSERIAL ANTIGENIC PEPTIDES
; FILE REFERENCE: 2300-1654 (PP01654.003)
; CURRENT APPLICATION NUMBER: US/10/111,983
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US-60/162616
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 37764
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 12967
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Neisseria
US-10-111-983-12967

```

```

Query Match          31.2%; Score 5; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      10 NKASS 14
Db      8 NKASS 12

```

```

RESULT 15
US-10-111-983-20596
; Sequence 20596, Application US/10111983
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: GALEOTTI Cesira
; APPLICANT: GRANDI Guido
; APPLICANT: MASIGNANI Vega
; APPLICANT: MORA Mariarosa
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: RAPPUOLI Rino
; APPLICANT: RATTI Giulio
; APPLICANT: SCARLATO Vincenzo
; APPLICANT: SCARSELLI Maria
; TITLE OF INVENTION: NEISSERIAL ANTIGENIC PEPTIDES
; FILE REFERENCE: 2300-1654 (PP01654.003)
; CURRENT APPLICATION NUMBER: US/10/111,983
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US-60/162616
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 37764
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 20596
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Neisseria
US-10-111-983-20596

```

```

Query Match          31.2%; Score 5; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      10 NKASS 14
Db      8 NKASS 12

```

Search completed: April 23, 2004, 15:05:25
Job time : 10 secs